Hearing this while seeing that: Semantic congruence affects processing of audiovisual stimuli

Von der Fakultät für Lebenswissenschaften
der Technischen Universität Carolo-Wilhelmina
zu Braunschweig
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)
genehmigte
D i s s e r t a t i o n

von Timo Martin Reißner
aus Braunschweig

1. Referent: Professor Dr. Dirk Vorberg
2. Referentin: Professor D. Brigitte Röder
eingereicht am: 08.10.2007
Mündliche Prüfung (Disputation) am: 16.04.2008
Druckjahr 2008

Veröffentlichung der Dissertation

Teilergebnisse dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Tagungsbeiträge:

Reißner, T. & Vorberg, D. (2006). Auditive Reize beeinflussen die Reaktion auf visuell eindeutige Ereignisse. In: Hecht, H., Berti, S., Meinhardt, G. & Gamer, M. (Hrsg.). *Beiträge zur 48. Tagung experimentell arbeitender Psychologen*, S. 181. Johannes Gutenberg Universität Mainz. Lengerich: Pabst Science Publishers.

Reißner, T. & Vorberg, D. (2007). Gackernde Hühner, klingelnde Gitarren und bellende Autos: Semantische Kongruenz bei audiovisuellen Reizen. In: Wender, K.F., Mecklenbräuker, S., Rey, G.D. & Weh, T. (Hrsg.). *Beiträge zur 49. Tagung experimentell arbeitender Psychologen*, S. 160. Universität Trier. Lengerich: Papst Science Publishers.

Contents

Summary

Under what circumstances does the human brain integrate multisensory information? Stein and Meredith (1993) have proposed three rules which describe the conditions of integration. Accordingly, temporal (1) and spatial (2) congruence facilitates behavioral responses and enhances neuronal firing. The weaker the stimuli, the larger is the enhancement [rule of inverse effectiveness (3)]. Are these rules sufficient to explain when stimuli from different modalities are integrated? In the literature, the rules were mostly tested with simple stimuli like flashes and clicks. But most audiovisual stimuli in the real world also contain semantic information. What happens if you hear something and see something else? Is semantic congruence irrelevant for multisensory integration? Or if semantics is relevant, are there any restrictions?

The present experiments investigate the role of semantic congruence in responding to audiovisual stimuli. Pictures and environmental sounds of animals and objects were presented in the first experiment. Participants were to categorize stimuli of a given target modality as living or nonliving. Results indicate that corresponding stimuli (e.g. a barking dog) elicited faster and more accurate responses compared to unimodal stimuli (e.g. picture of a dog without a sound), to incongruent stimuli (e.g. picture of a dog and the sound of a piano), and even to stimuli from the same category (e.g. a meowing dog). Thus, the results show clear effects of semantic congruence on processing audiovisual stimuli. These effects are not explainable by response-congruence.

The experiments included several kinds of stimuli and tasks to explore the generalizability of effects of semantic relation. Besides varying semantic congruence between pictures and environmental sounds, congruence of movement directions of simple visual and auditory stimuli, as well as between written and spoken words was varied. Tasks ranged from categorization, over detection tasks to reports of perception. Taken together, clear effects of semantic relation were found in tasks requiring processing of the content and in all kinds of employed stimuli. Vision dominated in most experiments, but effects from audition onto vision were also evident.

1. Introduction

We live in a multimodal world. All senses always gather information and our brain merges theses inputs into a coherent percept. For example, when visiting a zoo, we do not just see big grey elephants. We also hear them trumpeting, smell their dung, may even feel their hard leather skin. However, we often speak of our visual sense only. "I've *seen* that movie!", "Will you *watch* the soccer game?", "*Look*, there's a train coming!" – But other senses influence visual perception. When we watch television and turn off the volume, it would not be the same as with all sounds. Eating a meal is another everyday example for multisensory integration. Visual and olfactory as well as somatosensory and gustatory information is integrated in a special taste area in the caudolateral orbitofrontal cortex (Purves et al., 2004). Thus, our different senses do not operate independently but instead cooperate extensively.

These examples illustrate the relevance of multisensory perception. Unfortunately, researchers have mostly explored the senses as if they were independent of each other. An exception is an early attempt by Todd (1912). He found a reduction of response times to bimodal in contrast to unimodal stimuli. Later, a nowadays famous phenomenon was discovered by Howard and Tempelton (1966): The "ventriloquism effect" indicates that the voice of a ventriloquist seems to originate from a puppet. Thus, the source of auditory information is biased towards a potential visual source. This effect is also evident in movie theaters where the sound seems to originate from the actor's mouth but actually comes from loudspeakers at the sides. The ventriloquism effect illustrates influences of information from one modality onto another modality. This domination of one modality and different types of multisensory integration will be covered in more detail in chapter 1.1.

Where in the brain does multisensory integration occur? And how is information integrated? Neuronal correlates of multisensory integration are discussed in chapter 1.2. Under what circumstances is information from different senses combined? Available theories and rules are discussed in chapter 1.3. Stein and Meredith (1993) have summarized three rules, i.e. a temporal rule, a spatial rule and a rule of inverse effectiveness. The goal of the present work was to find out if these rules are sufficient to explain all phenomena or if an expansion

is needed. Specifically, the experiments explored whether semantic congruence is necessary for integration. Specifically, how does our brain process seeing an elephant and hearing a lion? This aspect has been widely neglected in the multisensory domain. Chapter 1.4 gives an overview of available studies on crossmodal effects of semantic relation, most of which have used linguistic stimuli. Speech perception has repeatedly been regarded as a special case of audiovisual integration (Tuomainen, Andersen, Tiippana & Sams, 2005). The main goal of the present experiments was to find out if information is integrated at an amodal semantic level when nonlinguistic stimuli from different senses are combined. The present experiments used different stimuli and different tasks to explore the premises, the level of integration (early vs. late) and the generalizability of crossmodal effects of semantic relation. Line drawings and environmental sounds of animals and nonliving objects, perceived movement directions of disks and tones as well as written and spoken words served as visual and auditory stimuli in different experiments. Implemented tasks were speeded categorization, detection tasks and reports of perception. The objectives of the experiments are discussed more precisely in chapter 1.5, followed by the experiments themselves and their discussions in the empirical part of this work.

1.1 Definition of multisensory integration

In the literature, several alternative terms, such as multisensory integration, intersensory perception, polymodal functions, amodal representations and crossmodal priming have been used to explain different aspects of the same phenomenon: the combination of information from different senses. Depending on objective, task and context of an experiment, the terms can have different meanings (Calvert, 2001). The present work mainly uses the term "multisensory integration" to underline the focus on the combination between the senses.
How is multisensory integration experienced? The experimental studies describing multisensory processing can be divided into two categories (Gondan, 2005). First, one response to a compound stimulus from two or more modalities is required. Thus, several senses are investigated together. Second, the effect of one modality onto another is explored. Participants respond to one modality (e.g. vision). Additionally, stimuli from another modality (e.g. audition) are presented, which may or may not influence processing of visual stimuli. This helps to investigate the influence of an irrelevant modality. Examples of these two categories are described separately in the following chapters.
Another way of categorizing studies of multisensory processing is the method of measuring multisensory integration. In other words, when do we know that crossmodal stimuli are integrated? A common measurement is the report of sensation. A famous example is the McGurk effect (McGurk & MacDonald, 1976). Observing a person whose lips form the syllable 'ga' while hearing the person say 'ba', most participants perceive 'da'. Thus, the subjective report is the dependent variable. Although the validity of subjective report is obvious, a disadvantage is its lack of objectiveness. Response times (RTs) are more objective and therefore also used frequently. Usually a response to two simultaneous stimuli from different modalities is faster than a response to stimuli from the same modality (Todd, 1912). Miller (1982) has found that responses to bimodal stimuli are even faster than would be predicted from a separate activation model. A separate activation model (or 'race model') assumes that two channels operate independently of each other, with the fastest channel determining the response time (RT). Conversely, when responses are faster than predicted

by a race model, coactivation is assumed. Systematic violation of the race model prediction is evidence for multisensory integration.
Diederich and Colonius (2004) have found that responses to trimodal stimuli (light, tone and vibration) are even faster than to bimodal stimuli and faster than predicted by a three-channel race model.
Within the last decades, functional imaging studies have become predominant. Functional magnetic resonance imaging (fMRI), positron emission tomography (PET) and event-related potentials (ERPs) shed light on where multisensory signals are integrated. Thus, integration can be distinguished from processing the modalities separately (Calvert & Thesen, 2004). Results of this approach are discussed in 1.2.
Other measures than subjective reports of perception, RTs and functional imaging, such as eye movements, have also revealed effects of multisensory integration (Kirchner & Colonius, 2005), but due to economy, theses studies play a subordinate role in this work.

1.1.1 Influences of one modality onto another

The majority of research on multisensory integration focuses on influences from one modality onto another, that is, how information of one modality affects perception of information from another modality.

Vision influencing auditory perception. The McGurk effect is an example for effects from vision onto audition. Specifically, seeing a speaker uttering ‘ga’ while hearing the spoken utterance ‘ba’ is mostly perceived as ‘da’ (McGurk & MacDonald, 1976).
The spatial location of auditory stimuli may also be affected by the location of visual stimuli. As mentioned earlier, the ventriloquist effect demonstrates how the perceived location of a sound is biased to the location of a light (Howard & Templeton, 1966).
Presentation of audiovisual stimuli can lead to a failure of responding to the auditory part. This so called Colavita effect (Colavita, 1974) describes the dominance of the visual part of a bimodal stimulus. Koppen and Spence (2007)

presented a visual, an auditory, or a visual and an auditory stimulus in each trial. Participants were to respond as fast and as accurately as possible to a visual stimulus with one key and to an auditory stimulus with another key. The results indicated that participants failed to respond to the auditory component of bimodal stimuli significantly more often than to the visual part.

Audition influencing visual perception. Within the last decade, several phenomena have been discovered that show influences from audition onto vision. For example, a sound may influence the perceived direction of a bistable visual motion display (Sekuler, Sekuler & Lau, 1997). When two identical objects (e.g. disks) move towards one another, coincide and then move away from each other, participants mostly perceive the disks as streaming through each other. The alternative perception (i.e. bouncing disks) is rarely seen (Metzger, 1934). However, presentation of a click simultaneously with the coincidence leads to severely increased bounce perceptions (Sekuler et al., 1997). This phenomenon is further discussed in Experiment 8. The two possible perceptions are illustrated in Figure 3.8.

Besides effects on bistable visual stimuli, audition can affect even unambiguous visual stimuli. A brief flash accompanied by two beeps is mostly perceived as two flashes. The illusory double flash persists even when participants are aware that only one flash was presented (Shams, Kamitani & Shimojo, 2000). Interestingly, a study using visual evoked potentials found almost identical potentials for the illusory flash and a physically double flash (Shams, Kamitani, Thompson & Shimojo, 2001). The brain does not seem do distinguish between actual visual perception and visual perception induced by auditory stimuli. Furthermore, effects of the sound are found even in the visual cortex: the two beeps influence predominantly visual areas and induce perception of two flashes.

1.1.2 Integrated Perception

Besides influences from one modality on another, the result of interacting modalities is often a common percept of several modalities. Synesthesia nicely describes this process. Persons with synesthesia see colors when they hear phonemes, or have tactile sensations when eating certain foods (Cytowic, 2002). Thus, a common sensation of different modalities is perceived, for example a green ‘a’.

An example for an integrated percept in everyday life is speech perception. Speech perception is significantly influenced by seeing and hearing a speaker, compared to just hearing speech. Especially when presenting noisy signals, the additional sight enhances speech perception (Sumby & Pollack, 1954). Thus, speech perception is commonly regarded as an integration of visual and auditory speech signals (Gondan, 2005). If so, the McGurk effect can also be regarded as an integrated percept. Visual and auditory information differs and is combined in a way that a new comprehensive percept arises (McGurk & MacDonald, 1976).

The debate of integrated perception versus modality dominance depends on the implemented task. When participants are instructed to indicate what they hear, researchers focus on an influence from vision onto audition. When the task is to respond to stimuli from all modalities, a common percept is often contrasted with a unimodal percept. Thus, focus of a study mainly determines whether the modalities merge to one percept or one modality influences another one.

For example, studies using a redundant target paradigm contrast responses to unimodal (e.g. visual or auditory) stimuli with responses to multimodal (visual and auditory) stimuli. The focus is set to the combination of both modalities, because this combination is compared to a single modality (cf. Diederich & Colonius, 2004). Todd already found in 1912 facilitated responses to multimodal compared to unimodal stimuli.

1.2 Neuronal correlates and mechanisms of multisensory integration

Where and how does the brain integrate information from the different senses? There is a tremendous amount of neuroimaging studies on multisensory integration. Several studies focused on the role of the superior colliculus (SC). The SC was originally considered exclusively as a site for visual properties, such as eye movements (Stein & Meredith, 1993). Early findings of multisensory neurons by Walker (1942, 1943; cited by Stein & Meredith, 1993) were not picked up by others. But from the 1960s on, the popularity of exploring the multisensory aspect of the deeper SC layers was rising (Stein & Meredith, 1993). In single cell animal studies the mechanisms of multisensory integration were explored (for a review see Stein, Jiang and Stanford, 2004). The underlying neuronal processes are discussed in 1.2.2. Before, the sites of multisensory integration are reviewed.

1.2.1 Sites of multisensory integration

When brain areas processing multisensory stimuli were first discovered, it was widely believed that information from different senses was processed hierarchically from the unisensory cortices (e.g. visual and auditory cortex) to the association cortex. Thereby, it had been hypothesized that integration occurred just in association areas. These areas were defined by receiving inputs from more than one modality and containing neurons that respond to multiple modalities (Calvert & Lewis, 2004). Such areas were located mostly in anterior and posterior portions of the superior temporal sulcus (STS) (e.g. Watanabe & Iwai, 1991; Desimone & Ungerleider, 1986), the parietal cortex (e.g. Bremmer et al., 2001) and frontal cortex (e.g. Watanabe, 1992; Fuster, Bodner & Kroger, 2000). As mentioned above, subcortical areas as, for example, the SC (Barraclough, Xiao, Baker, Oram & Perrett, 2005), the caudate nucleus and the substantia nigra (Nagy, Eördegh, Paróczy, Márkus & Benedek, 2006) also participated in multisensory integration.

The assumption of hierarchical processing implies a model of late multisensory processing. On the other hand, there are several accounts that argue for an early model (Calvert Thesen, 2004). First of all, audio-visual interactions were found in the alleged unimodal visual areas. Macaluso, Frith and Driver (2000) found enhanced activity in the visual cortex to spatially aligned visual-tactile stimuli. Second, just 40 ms after stimulus onset, interactions between vision and audition were observed in the visual cortex (Giard & Peronnet, 1999). Thus, a pure feedback account seems unlikely. The 'early' vs. 'late' debate will be further discussed in 1.2.2 when aspects of neuronal processing (i.e. feedforward vs. feedback projections) are explained.

Up to this point, several areas were listed that process multisensory information. Now the question arises what the purpose of the different areas is. Calvert (2001) gives an overview of which area processes what kind of information. Accordingly, the STS appears important for integrating featural information, similar to the 'what' path in visual perception (Ungerleider & Mishkin, 1982). This field is mostly explored in audiovisual speech signals (cf. 1.4.1). Multisensory spatial cues (similar to the 'where' path) are mainly mediated in the SC and the intraparietal sulcus (IPS). Temporal information is integrated also in the SC and additionally in the insula, both of which are subcortical areas. The role of the frontal cortex is less clear, but this region seems to be involved in integrating newly acquired associations. According to Calvert and Thesen (2004), the outcome of the processes in the associative areas is sent via feedback projections to the presumed unimodal areas.

Although several explored the role of the multisensory regions, the overwhelming plasticity of the human brain shows that these functions have not yet been entirely determined. For example, congenitally blind persons were PET-scanned by Ptito, Moesgaard, Gjedde and Kupers (2005) before and after a training of detecting the letter 'T' on an electrode array with their tongue. Before training the groups did not differ, but afterwards increased activity was found in the occipital cortex of blind but not of sighted participants. Two conclusions can be drawn. First, training of seven hours changed the cortical activation patterns of blind participants. Second, a somatosensory task led to activation in the visual cortex. Both conclusions underlie the plasticity of the brain. This is supported by increased activity of the primary visual cortex in blind

participants during Braille reading, as compared to normal controls (Sadato et al., 1996). Furthermore, animals whose projections were rewired so that auditory information was processed in the visual cortex, interpret auditory stimuli as if they were in fact visual (Sur, 2004).

The site of integration is also modulated by attentional mechanisms. Johnson and Zatorre (2005) found increased activity in multimodal and unimodal areas of the target modality, whereas activity in the cortices of the unattended modality was decreased. Thus, multisensory integration also depends on top-down influences.

To sum it up, several sites participate in integrating multisensory information. Integrating neurons were found in earlier (predominantly unimodal) as well as later (associative) areas. For example, audiovisual stimuli are processed in visual and auditory cortices as well as in multisensory areas. The specific area depends on multiple conditions, such as employed stimuli and tasks. Figure 1.1 gives an overview on integration sites. The connections between the participating areas are discussed in the next section.

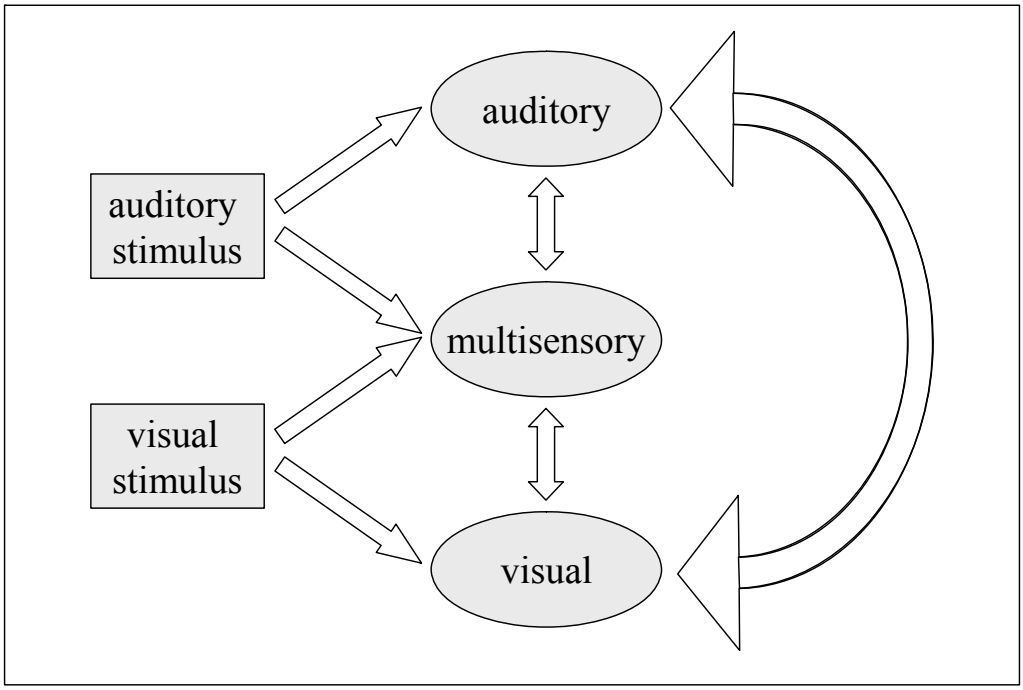


Figure 1.1: Model of neuronal connections for multisensory integration. Stimuli project directly to their unimodal regions as well as multimodal regions. Unimodal and multimodal regions are connected bidirectionally via feedforward, feedback and lateral connections.

1.2.2 Neuronal mechanisms of multisensory integration

How does the brain integrate information from different senses? Section 1.2.1 described brain areas whose activity was increased during multisensory tasks compared to unimodal tasks. This result did not essentially imply combination at a neuronal level (Calvert, 2001). An increasing amount of active neurons that interact with one another may also explain the result. However, neurons that respond to more than one modality have been found in several multisensory areas (Desimone & Gross, 1979). These cells have largely overlapping receptive fields. For example, Meredith and Stein (1996) found 86% overlapping receptive fields of neurons responding to auditory and visual stimuli in the cat SC. The modalities are organized in a map-like fashion in these multisensory neurons (Wallace, Wilkinson & Stein, 1996). Thus, the modality specific information is spatially aligned. On a neuronal basis, multisensory integration is evident when the firing rate of bimodal stimuli exceeds the sum of both unimodal stimuli (Meredith & Stein, 1983). This supra-additive response enhancement is not found in unimodal integration and is thus a good indication of multisensory integration (Alvarado, Vaughan, Stanford & Stein, 2007). Stimuli that are minimal effective produce the largest response enhancement. This effect is called inverse effectiveness. For example, Meredith and Stein (1986) have found neurons that produced response enhancement of 1207% with minimal effective audiovisual stimuli compared to the most effective unimodal response. On the other hand, when bimodal stimuli are not spatially aligned, responses can be depressed compared to the unimodal stimuli alone (Stein, Jiang & Stanford, 2004). The neuronal processes behind these sub-additive effects can be described by assuming that a multisensory cell has input from two modalities. Input from one modality is thereby excitatory while the input from another modality is inhibitory. Thus, depression of activation arises (Meredith, 2002).

The neuronal system also takes reliability of the stimuli into account (see also 1.3 for the rules of multisensory integration). Visual and auditory stimuli are combined in a statistical optimal way (Ernst & Banks, 2002). Banks (2004) has described that the best combination is a weighted average of two stimuli. Costs

and benefits are thus considered by Bayes' rule which always results in more reliability than either stimulus alone. A model using maximum-likelihood to integrate visual-haptic stimuli behaved very similarly to human participants (Ernst & Banks, 2002). Specifically, when ambiguous visual stimuli are presented with clear auditory stimuli, the neuronal system compensates for the larger variance of the visual stimuli. As a consequence, auditory stimuli dominate perception.

After exploring the processes within a neuron, I would now like to shed light on processes between neurons. How are the different integration areas connected? In the multisensory domain, this mainly comprises the debate of feedback versus feedforward connections between higher-order multisensory areas and lower-order presumably unimodal areas. As mentioned in 1.2.1, until recently it has widely been believed that integration solely occurs in the multisensory sites (e.g. the STS and the SC). When integration in unimodal areas was first observed, this was explained by assuming feedback connections from multisensory to unimodal areas (Calvert & Thesen, 2004). However, several findings argue against this hypothesis. Giard and Peronnet (1999) discovered effects of multisensory compared to unimodal stimuli 40 ms after stimulus presentation in V1. Differences in multimodal areas arose approximately 150 ms after stimulus presentation. Thus, the differences in V1 cannot be explained by feedback projections. Murray et al. (2005) discovered activities after 50 ms that depended on spatial alignment of the stimuli. Consequently, the spatial rule described above is evident in early areas. Furthermore, connections between early sensory regions were found (e.g. between V1 and auditory cortex), which speaks for lateral connections. These results may explain how the multisensory stimuli are bound in the early areas (Foxe & Schroeder, 2005). Multisensory stimuli are integrated even if higher multisensory regions are lesioned (Ettlinger & Wilson, 1990).

Driver and Spence (2000) have suggested a model that includes bidirectional connections between multimodal and unimodal areas as well as between unimodal areas. Thus, feedforward, feedback and lateral projections are embedded. Neuronal projections in multisensory processing were summarized and illustrated in Figure 1.1.

1.3 Rules of multisensory processing

Under what circumstances does the brain integrate information from different senses? Not every piece of information can be integrated with every other piece. A plausible example is when you hear somebody walking but all you can see is a sitting man, these two stimuli are unlikely to be integrated.
Three simple rules have first been mentioned by Stein and Meredith (1993) and are now widely accepted (e.g. Holmes & Spence, 2005; Giard & Peronnet, 1999). The rules mainly base on the previously mentioned studies of the SC. Hence, they explain under what circumstances the neuronal response to multisensory stimuli is enhanced compared to the sum of the unimodal stimuli. The spatial rule, the temporal rule and the inverse effectiveness rule are described below.

According to the 'spatial rule', multisensory stimuli are integrated when they originate from approximately the same location. When a light and a tone was presented at the same location, responses in SC of cats were enhanced compared to the sum of unimodal stimuli. Accordingly, the cats detected food underneath the stimuli more quickly and accurate. When light and tone mismatched, responses were depressed and correct detections declined. The further the stimuli were apart, the lower the response rate (Stein, Meredith, Huneycutt & McDade, 1989). In human participants, responses are facilitated and detectability is enhanced to spatially concordant audiovisual stimuli (McDonald, Teder-Sälejärvi & Hillyard, 2000). On the first view, these results challenged the ventriloquist effect, which showed that sight and sound do not need to originate from the same location to get integrated by the brain. However, Alais and Burr (2004) varied spatial correspondence of light and sound as in the previous experiment. Additionally, they blurred visual stimuli, which resulted in less accuracy. Again, for unambiguous stimuli, localizations shifted towards the visual stimuli. In the blurred condition, however, audition dominates over vision. Thus, reliability of the stimuli influences how the modalities are integrated. In the case of spatial resolution, vision is about 5-10 times more accurate than audition (Banks, 2004). Resuming to the ventriloquist effect, the auditory stimulus (i.e. the speaker's voice) seems to originate from the visual sti-

mulus (i.e. the puppet), because vision dominates audition in spatial tasks. Interestingly, the McGurk effect (McGurk & MacDonald, 1976) is also unaffected by spatial disparity (Spence & Driver, 2004). An exception to the spatial rule is described by Gray, Tan and Young (2002). They found that vibration in the back improves detectability of visual changes in the front more than a tone. Consequently, haptic stimuli seem to be integrated with visual stimuli even if they are spatially apart.
The neural mechanisms behind the spatial rule rely on overlapping receptive fields of the multisensory cells for visual and auditory stimuli (Stein, Jiang & Stanford, 2004).

The 'temporal rule' states that if multisensory stimuli occur at about the same time, responses are enhanced (Holmes & Spence, 2005). More precisely, the activity patterns of two stimuli need to overlap (Stein & Meredith, 1993). Thus, a time window exists in which stimuli are integrated due to their temporal alignment.
The time windows must be highly flexible, because stimuli may originate from the same object, but neural responses are elicited at different times. For example, as light travels faster than sound, sounds of objects that are further away arrive later than visual components of the same object. Arnold, Johnston and Nishida (2005) tested this experimentally by varying the distance of stimuli (i.e. stream-bounce display with or without a sound; cf. Experiment 8) between approximately 1 to 15 meters. Alternatively, the sound was presented via headphones. They found an increasing effective time window (from audition onto vision) with increasing stimulus distance. However, this effect did not arise when the sound was presented via headphones. Thus, the system seems to adjust the effective time window based on the delay between sight and sound. This adjustment depends on previous experience, which is indicated by results of Fujisaki, Shimojo, Kashino and Nishida (2004). In an adaptation phase they presented a flash and a tone separated by a fixed stimulus-onset asynchrony (SOA). In the following test phase with different stimulus-onset-asynchronies (SOAs), they found a shift in simultaneity judgments to the SOA of the adaptation phase.

Such time windows also exist in unimodal integration. In the visual domain, two dimensions (i.e. color and form) are integrated if they both occur in a window of 40 ms. Interestingly, in multisensory tasks this window may be as large as 600 ms. For example, the McGurk effect persists within a window of 180 ms (Calvert & Thesen, 2004).

Auditory stimuli are usually more reliable than visual stimuli in the temporal domain, whereas visual stimuli dominate spatial tasks (Welch & Warren, 1980). However, the effect of temporal alignment also depends on specific stimulus features. With increasing uncertainty of the auditory stimulus, vision dominates audition. Heron, Whitaker and McGraw (2004) increased the uncertainty of the location of a visual dot by increasing its size. They found increased effects of an auditory stimulus on locating the dot, when the dot was enlarged. When increasing the uncertainty of the auditory stimulus, the effect on locating the dot decreased. Hence, reliability affects domination of one modality.

The third rule is called ‘inverse effectiveness’. The weaker two stimuli, the larger is the response enhancement of their integration (Stein & Meredith, 1993). Meredith and Stein (1986) found more enhancements of responses to audiovisual stimuli compared to the sum of the responses to visual and auditory stimuli alone, with decreasing stimulus intensity. Specifically, with optimal effective stimuli, the enhancement was 110%, with sub-optimal enhancement 258%, and with minimal effective stimuli the enhancement was 483%. This superadditivity is also evident in enhanced detection of multisensory stimuli in human participants (Bolognini, Rasi & Làdavas, 2005).

Why is the firing rate of neurons increased beyond the sum of the responses to the unimodal stimuli? And why are additional spikes elicited? Stanford, Quessy and Stein (2005) suggested an answer to these questions. They varied stimulus intensities of audiovisual stimuli while measuring extracellular activity activities in cat’s SC neurons. The results support the conclusion that the lower the stimulus intensity, the higher the response rate. Conversely to studies measuring firing rates, they found purely additive postsynaptic potentials due to multisensory stimuli. Consequently, superadditive firing rates reflect temporal summation of postsynaptic potentials. Subadditive firing rates are caused by

the lack of summation. However, such a linear model is insufficient to explain all findings, which necessitates a more complicated model (Holmes & Spence, 2005).

Schnupp, Dawe and Pollack (2005) have quantified the rules by assuming a psychophysical modal and taking Euclidean Distance between multisensory stimuli into account. Based on these assumptions, predictions of stimulus detection were enhanced.

The three rules mainly base on studies of the monkey's and cat's SC. In humans other factors may influence multisensory integration as well. For example, Calvert, Hansen, Liu, Lloyd and McGlone (2001) found that task demands and the target modality affected responses.

1.4 The role of semantic congruence in multisensory integration

Besides the three rules mentioned above, semantic congruence is believed to be a critical factor for multisensory integration (Calvert & Thesen, 2004). Research on this topic has mostly been performed with linguistic stimuli. Therefore, this research area is subsequently reviewed, and it is discussed whether speech is a special case of multisensory processing (Calvert et al., 2004). The few non-linguistic studies are discussed afterwards.

1.4.1 Audio-visual speech perception

Several experiments about semantic congruence of multisensory stimuli have been carried out with speech stimuli. Calvert, Campbell and Brammer (2000) presented auditory speech signals and videos of lip movements to participants while using fMRI to scan their blood oxygen level dependency (BOLD) signals. The stimuli were congruent (story was heard while the corresponding lip movement was seen) or incongruent (different stories in visual and auditory streams). Furthermore, the authors included unimodal speech signals by switching off sound or sight for unpredictable time lengths. The main finding was an increased BOLD signal (supra-additive compared to the unimodal stimuli) in portions the left superior temporal sulcus (STS) during congruent audio-visual speech, while the BOLD signal was reduced in this area during incongruent signals (sub-additive compared to the unimodal stimuli). In addition, with similar congruent and incongruent audiovisual stimuli, Calvert et al. (1999) found enhanced activation in auditory and visual cortices. In this study, incongruent stimuli were semantically incongruent but also temporally incongruent, as the lip movement never matched the spoken words, i.e. the mouth could be opened while no sound came out.

As evident in processing of other stimuli, activated brain areas largely depend on the implemented task. Thus, a specific determine area cannot be defined (Campbell, 1998).

In the classic version of McGurk and MacDonald (1976) a visual presented speaker moves his lip for saying the syllable /ga/ whereas auditory /ba/ is presented. Most of the time, the incongruent information is perceived as /da/. This effect is also present when presenting only isolated kinematic properties of the moving lips (light points that were attached to the moving face), without conscious knowledge of the participants what the visual stimulus represents (Rosenblum & Saldaña, 1996). Besides the lips and mouth movements, seeing facial expressions without the mouth also improves identification of spoken words (Munhall & Vatikiotis-Bateson, 1998).

Written and spoken words are also common in multisensory tasks. Most of them focus on crossmodal priming. Specifically, a stimulus in one modality should enhance processing of another stimulus in a different modality.
Holcomb and Anderson (1993) have found semantic priming effects from audition onto vision and vice versa. The experiment was set up by a written word (e.g. salt) and a spoken word (e.g. pepper) that were separated by a SOA of 0 ms, 200 ms or 800 ms. A speeded lexical decision task resulted in enhancement of semantically primed targets. This study is further discussed in section 4, when follow-up experiments were conducted.

There is a long debate whether linguistic stimuli are processed differently from nonlinguistic stimuli (e.g. Hauser, Chomsky & Fitch, 2002). For example, Tervaniemi and Hugdahl (2003) proposed that auditory speech signals and non-speech signals are processed by different brain mechanisms. Tuomainen, Andersen, Tiippana and Sams (2005) explored if participants process auditory signals differently, whether or not they know the signals contain speech. They presented masked auditory speech signals throughout the experiment but varied instructions. In one condition, participants were to categorize stimuli into non-speech categories, whereas in another condition, they knew that the signals contained speech. Additionally, mouth movements were presented visually. In unimodal auditory as well as congruent audiovisual conditions, correct responses did not differ according to the instruction. However, large differences were found for incongruent audiovisual signals. Responses were much more accurate without knowledge of contained speech (84%) compared to responses

with the knowledge that the same stimuli contained speech (29%). Presentation of unmasked speech led to an even larger decrease of correct responses in the incongruent condition (3%). These results show the identification of speech has a large impact on how auditory stimuli are processed.

Visual speech signals are processed very similarly as auditory speech signals. Specifically, observing a face that makes speechlike movements activates the auditory cortex. In contrast, nonspeech movements do not engage in increased activation of the auditory cortex. Furthermore, activation of the auditory cortex states early integration mechanisms (Calvert et al., 1997).

Does the activation of the auditory cortex imply speech integration at an early level? Musacchia, Sams, Nicol and Kraus (2006) found integration of lip-reading and auditory speech signals about 11 ms after the arrival of acoustic signals in the ear. This supports an approach of early integration.

1.4.2 Semantic influences in nonlinguistic stimuli

Only a few studies have been carried out on the effects of semantics with non-linguistic multisensory stimuli.

Laurienti, Kraft, Maldjian, Burdette and Wallace (2004) combined linguistic and non-linguistic stimuli. They presented a blue or a red disk. In unimodal trials, the word ‘red’ or ‘blue’ was also presented visually, while in crossmodal trials, the words were presented auditory. Participants’ task was to indicate the color of the disk. Responses in crossmodal trials were faster than those in unimodal trials, thus supporting larger influences of crossmodal cues than of unimodal cues.

Molholm, Ritter, Javitt and Foxe (2004) presented pictures and environmental sounds of animals, and varied their congruence. The task was to detect a target animal in any modality. For visual and auditory stimuli corresponding to the same animal, responses were faster and more accurate than for different animals. Yuval-Greenberg and Deouell (2007) also presented pictures and sounds of animals, but used slightly different instruction and task. Participants were to attend to either the visual or the auditory modality, in contrast to attending to both modalities in Molholm et al. (2004). After presentation of each audiovi-

sual pair a forced-choice question was presented (e.g. ’dog?’). If the target animal was present, the appropriate key was to be pressed, if not a different key was to be pressed. Results showed that responses were faster and more accurate in congruent than in incongruent trials. The influence from vision onto audition was larger than the opposite effect. Simultaneously with the behavioral task, EEG was measured. Induced Gamma-band activity was found to be enhanced about 260 ms after stimulus onset for congruent as compared to incongruent trials. Multisensory trials differed from unimodal trials already 90 ms post-stimulus-onset, but the semantic relationship had no effect yet. This result indicates that semantic integration occurred at a later processing stage.

The studies of Molholm et al. (2004) and of Yuval-Greenberg and Deouell (2007) both supported the role of semantic congruence. However, semantically congruent stimuli were also response-congruent in these experiments. For example, picture and sound of a cow should elicit the same response, that is, ‘yes, the target was present’. This condition was compared to the presentation of the picture of a cow and the sound of a dog. Here, the correct response is again ‘yes’, but the response to the dog would be ‘no’. Thus, semantically incongruent stimuli were also response-incongruent. This hypothesis may be ruled out by comparing ‘no’ responses, because then same and different stimuli were both response-congruent. Molholm et al. (2004) found in a non-hypothesized post-hoc analysis a difference between these conditions in the N400. The N400 component is believed to reflect semantic processes (Kutas & Federmeier, 2000). Behavioral differences were not reported.

Congruence effects were also found with more abstract stimuli. Perception of an ambiguous drifting grating is effected by the simultaneous presentation of a falling or rising tone (i.e. its pitch is descending or ascending) (Maeda, Kanai & Shimojo, 2004). Participants reported that they perceived the grating as if moving in the same direction as the tone. Whether or not this effect is caused by semantic congruence is further discussed in experiments 5 through 7.

A few studies about semantic congruence of multisensory stimuli have also been carried out in other domains. For example, memories for congruent multisensory stimuli are better than for incongruent or unimodal stimuli, and activated sites differ (Murray, Foxe and Wylie, 2005). Effects of semantics have

also been observed with visual-haptic stimuli (Helbig & Ernst, 2007) and visual-olfactory stimuli (Gottfried & Dolan, 2003).

Thus, there are some studies that found that semantic congruence enhances multisensory integration. However, in other studies such effects were absent. E.g., Koppen, Alsius and Spence (2008) studied the role of semantic congruence in the Colavita effect (see 1.1.1). No difference was found for semantically congruent stimuli (i.e. sight and sound of a dog) compared to incongruent stimuli (i.e. sight of a dog, sound of a cat). This absence of an effect might be caused by low level processing. On the other hand, spatial and temporal coincidence influence the Colavita effect (Koppen & Spence, 2007).
The role of semantic congruence for audiovisual integration remains unclear. If semantic congruence is important, it remains unclear under which circumstances effects of semantic relation can be found. And at what processing stage might they occur?

1.5 The present experiments

The three rules of multisensory integration have been explored extensively. Since Stein and Meredith (1993), it has been widely accepted that temporal and spatial concordance as well as inverse effectiveness affect processing of multisensory stimuli. Instead of questioning the rules and exploring their effects, I speculate whether any other stimulus properties influence multisensory processing. Are the rules sufficient to explain the conditions of multisensory integration? Under which circumstances are the modalities processed separately?

My hypothesis was that the semantics of a piece of information enhances integration. Calvert and Thesen (2004) already mentioned that semantic congruence may influence multisensory integration. They described in a real world example that a dog's bark and the sight of a cat can barely be integrated. Can this be supported by behavioral experiments? In linguistic research, Holcomb and Anderson (1993) have shown crossmodal effects of semantic relation. The role of semantics in nonlinguistic stimuli remains unclear because previous results may also be explained by response-congruence (cf. 1.4.2).

Thus, the present thesis focuses on effects of semantic congruence in nonlinguistic stimuli. *Experiment 1* explores effects of semantic congruence by varying congruence of pictures and sounds in a living/nonliving categorization task. It is hypothesized that the same picture and sound elicit faster and more accurate responses than response-congruent, incongruent or unimodal responses. By defining a target modality (i.e. the to-be-attended modality) a dominant direction may be explored. The results of the response-congruent condition (e.g. cow and dog) are subdivided in two groups. Semantically near and semantically far combinations will be determined and contrasted. The question is if a dog and a cat elicit faster responses than a dog and a snake. Furthermore, I investigate temporal aspects of integration by varying SOA in *Experiment 2*. The objective is to find out if semantic effects are larger at simultaneous or at successive presentation. The role of stimulus reliability will also be explored. Finding out when stimuli are perceived as simultaneous is a goal of *Experiment 4*. Here, a temporal order judgment task will be implemented. Additionally, effects of semantics are explored in this low level task. The question is at which

processing level integration occurs. When semantics has an effect in a low level task, deep processing is not needed. Thus, integration is likely to occur at an early level. *Experiment 3* has the same focus but uses a different task. Participants will be instructed to respond with a key press to visual and auditory presentation. No response is necessary when just one modality was presented (go/no-go task). This task does thus not require processing of the content.

Besides exploring effects of semantics in pictures and sounds of animals and objects, more abstract stimuli will be introduced to generalize the findings. In contrast to the previous experiments, dynamic visual stimuli will be used. The reason is that static pictures can hardly elicit dynamic sounds. Maeda et al. (2004) have found that movement of pitch affects perception of movement of an ambiguous grating. *Experiment 5* thus varies movement direction of an unambiguous disk along with movement of pitch in a tone. It is again predicted that responses to congruent stimuli are facilitated. *Experiment 6* controls in which direction auditory stimuli are perceived to move. Different movement directions of the visual stimuli will be introduced accordingly. In *Experiment 7*, I investigate the role of the processing level by implementing a simpler task. Participants are instructed to give a speeded response to any auditory stimulus. In contrast to Experiment 5 and 6, a task more elementary than categorization is used.
Semantic effects on the perceived direction of the motion path of two disks will be explored in *Experiment 8*. Expanding results of the stream-bounce paradigm (Sekuler et al., 1997), I present ambiguous as well as non-ambiguous paths of two disks (i.e. stream vs. bounce). Instead of presenting a meaningless, realistic sounds of streams and bounces are presented. The goal is to find out whether semantically different sounds induce perception of streaming or bouncing disks.

One advantage of linguistic as compared to nonlinguisitc stimuli is that semantic congruence can be varied in smaller steps with words than with pictures and sounds. Instead of stimulus-congruence (e.g. a trumpeting elephant) in nonlinguistic stimuli, semantic relation can be varied more precisely with linguistic stimuli. For example, the words ‘mother’ and ‘father’ are semantically related,

whereas ‘calf’ and ‘father’ are rather unrelated. Such a variation can hardly be achieved with nonlinguistic stimuli. *Experiment 9* explores in a categorization task of written and spoken words if semantic congruence influences responses. *Experiment 10* searches for effects of semantic relation in incongruent stimuli. For example, the living item ‘horse’ is response-incongruent to the nonliving item ‘saddle’. However, they are semantically related, whereas the prime ‘bottle’ is unrelated and response-incongruent. Furthermore, repetition priming or stimulus-congruence will be used. Response-congruent stimuli should elicit faster responses in these linguistic experiments than in the experiments using nonlinguistic stimuli, because of a high correspondence of the orthography in written words and the phonology in spoken words.

2. Semantic congruence of environmental sounds and pictures

2.1 Experiment 1: Semantic congruence vs. response-congruence

According to Stein and Meredith (1993), temporal and spatially congruent crossmodal stimuli tend to be integrated. The role of semantic congruence remains unclear. Some studies have focused on its role in linguistic stimuli. For example, Holcomb and Anderson (1993) found semantic influences in a crossmodal priming paradigm. Laurienti et al. (2004) combined words and nonlinguistic stimuli. The spoken color 'red' enhanced detection of a red disk stronger than the written word 'red'.

Up to now, little research has been performed with nonlinguistic stimuli. Lehmann and Murray (2005) contrasted visual and audiovisual stimuli in a memory recognition task. They used line drawings of different objects as visual stimuli. Auditory stimuli were environmental sounds of the same objects. Participants were to categorize objects as old (seen before) or new (not seen before). The initial presentation contained unimodal visual items in 50% of the trials and congruent as well as incongruent audiovisual items in 25% each. All repeated presentations were unimodal visual. The authors found more correct responses for repeated presentations that were initially congruent and bimodal, compared to initially incongruent and unimodal presentations. RTs did not differ for repeated presentations. Thus, a single presentation of a congruent sound along with a picture enhances memory for that picture.

What is the role of semantic congruence for perception? Are congruent audiovisual stimuli perceived and processed faster and more accurately than incongruent or unimodal stimuli? Molholm et al. (2004) as well as Yuval-Greenberg and Deouell (2007) have shown congruence effects in detecting pictures and sounds of animals. It remains unclear though, whether facilitation was due to semantic congruence or to response-congruence. Therefore, a design is needed that varies congruence levels more finely, i.e. from incongruence, over response-congruence (different stimuli emerging the same response) to stimulus-congruence (same item).

Consequently, nonlinguistic stimuli were employed in Experiment 1. Living and nonliving items were presented visually as line drawings. In the auditory modality, a compatible sound was presented simultaneously with the picture. The task was to categorize the stimuli in one predetermined modality (visual or auditory) as living or nonliving. Congruence between visual and auditory stimuli was varied. It was varied whether stimuli originated from the same object (i.e. picture and sound of a chicken), from different stimuli within the same category (i.e. picture of a chicken and sound of a cat), from different categories (i.e. picture of a chicken and sound of a car). In a unimodal condition, a stimulus from one modality was presented alone (i.e. either a picture or a sound). Participants responded to visual stimuli in one session and to auditory stimuli in another session, which allows to explore influences from vision onto audition and audition onto vision.

Furthermore, differences within the response-congruent condition were explored. The question was whether semantically near stimuli (e.g. sound of a cat and picture of a dog) elicit faster responses than semantically far stimuli (sound of a cat and picture of a snake). Thus, the response-congruent condition was subdivided into semantically near and semantically far. To generate the subcategories, a group of naive participants generated subcategories of the words of all stimuli. This categorization was done separately for living and nonliving categories. Words instead of pictures or sounds were used to ensure an amodal representation. A cluster analysis revealed overall subcategories, which were used to find semantically near and far conditions. Differences were explored with a post-hoc analysis.

The main hypothesis of Experiment 1 was that increasing congruence facilitates responses and decreases errors. Specifically, stimulus-congruent stimuli should elicit the fastest responses, followed by responses in the response-congruent condition. RTs in both conditions should be faster than in the unimodal condition, which indicates a response enhancement of the irrelevant crossmodal stimuli. A difference between the stimulus-congruent and the response-congruent condition would indicate semantic influences not explainable by response-congruence. Incongruent stimuli should lead to inhibit responses and result in the slowest responses. Stronger effects are expected for the audi-

tory targets because auditory stimuli are harder to identify and therefore giving visual stimuli an advantage.

2.1.1 Methods

Participants. Eight undergraduate students (2 male) received course credit for participation. The average age was 22.9 years (range from 19 to 38). They reported normal or corrected-to-normal vision, as well as normal hearing. All participants reported right-handedness. All were naive as to the purposes and hypotheses that motivated the study.

Apparatus. In all present experiments, stimuli were presented on a Microsoft Windows XP® operated PC with 1.0 GHz and 256 MB RAM. Background processes such us networking were shut down if possible to eliminate influences on stimulus presentation and processing of response inputs. Visual stimuli were presented on a 17'' Iiyama® CRT monitor with a refresh rate of 75 Hz. Auditory stimuli were presented via Creative SB Audigy 2 ZS® sound card and Yamaha stereo headphones. The experiments were controlled by MATLAB® 7 and Psychophysics Toolbox 2.54 (Brainard, 1997; Pelli, 1997).
Participants sat at a distance of 62.5 cm from the monitor and responded on a standard computer keyboard. A chin rest restricted head movements. The room was dimly illuminated.

Stimuli. Thirty-six pictures and 36 sounds were used. The stimuli from each modality consisted of 18 living and 18 nonliving stimuli. Stimuli classified as living were all animals. Other living items, such as human body parts, professions and family members, which are common in unimodal visual categorization tasks (e.g. Barbarotto, Laiacona, Macchi & Capitani, 2002), were omitted due to difficulties in presenting them as auditory stimuli. Musical instruments, vehicles, and objects of daily use were classified as nonliving.
The visual stimuli were grayscaled drawings created by Rossion and Pourtois (2004), which are similar to the picture set of Snodgrass and Vanderwart (1980). The stimuli were presented for 500 ms in a size of 300 by 300 pixels

(8.2 x 8.2°) in the center of a white screen. A list of the stimuli is given in the appendix.

The auditory stimuli were taken from several free sources (internet sites and CDs) and modified so that they were wav-files in mono with a sampling rate of 44100 Hz and 16 bit. After transformation, all sounds had approximately the same volume and a duration of 500 ms. Sounds were presented via Sennheiser headphones. Waveforms of all auditory stimuli are shown in the appendix.

In a pretest, I assessed identification and classification rates for auditory and visual stimuli with eight naive participants (one male; mean age: 22.8 years). After presentation of each stimulus, participants were instructed to name the presented stimulus, and to classify it as living or nonliving. Each auditory stimulus was presented once in the first block and each visual stimulus once in the second block to assure no benefit from the visual stimuli. Rates of correctly identified and classified stimuli are summarized in Table 2.1.

Table 2.1: Rates of correct identification and classification in percent for auditory and visual stimuli as a function of living and nonliving categories. Data were surveyed in a pretest of eight participants.

	identification		classification		overall
	auditory	visual	auditory	visual	
living	68.1	84.0	94.4	97.9	86.1
nonliving	52.1	79.2	77.1	82.6	72.8
overall	60.1	81.6	85.8	90.3	79.4

Design. The experiment used a 2 x 4 repeated measures design with factors *Target Modality* and *Congruence Type*. Target Modality indicates the modality which participants attended to. In two separate sessions, participants attended either to the visual or the auditory modality, while the other modality was irrelevant. The order was balanced over participants. Levels of Congruence Type were stimulus-congruent, response-congruent, incongruent and unimodal). An example for each level is shown in Figure 2.1. RTs and error rates served as dependent variables.

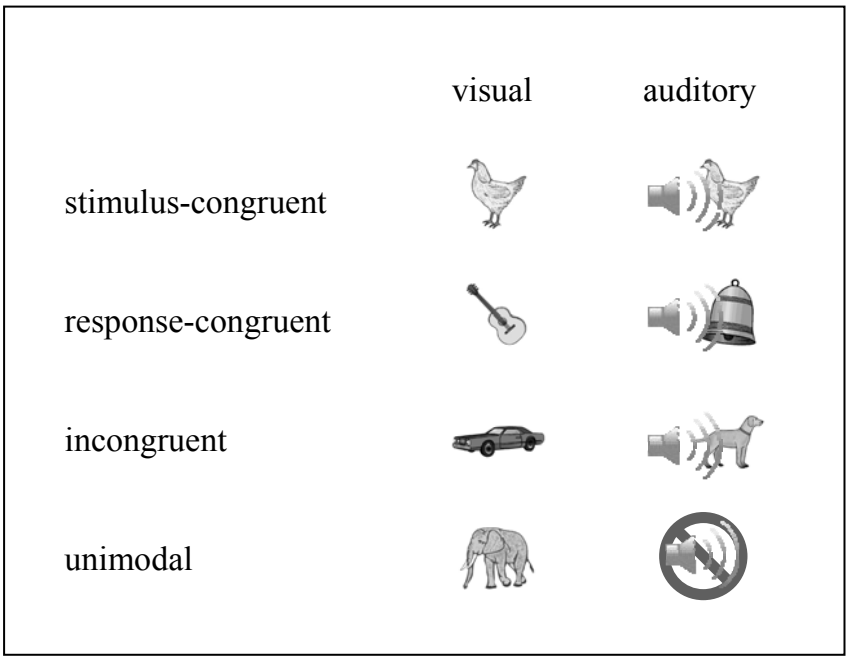


Figure 2.1: Congruence Type levels of Experiment 1. Target Modality is visual in this example, because no sound is presented in the unimodal condition.

Task. The task was to classify the stimulus in the target modality (visual or auditory) as living or nonliving. Participants were instructed to respond to the target modality, while the other modality was irrelevant and could be ignored. They were to permanently watch the monitor and to not remove the headphones. Further they were to respond as quickly and accurately as possible.

Procedure. The experiment was conducted in two separate sessions, mostly on consecutive days. Half of the participants responded to the visual stimuli in session one and to the auditory stimuli in session two, and vice versa for the other half of participants. Instructions were given onscreen; participants were to give a short summary of the task to assure understanding. Each session included 720 trials and took approximately 50 min. for completion.

Figure 2.2 shows the trial events on an example trial. A trial started with a black fixation cross (size: 0.3° x 0.3°) in the center of the screen for 1000 ms. Then an auditory and a visual stimulus were presented for 500 ms. With the onset of the stimuli, RTs were measured, and participants categorized the target stimulus by pressing the left or right control keys. Errors were signaled by a

black X (height: 1.4°) for 500 ms. After 1000 ms the next trial started automatically.

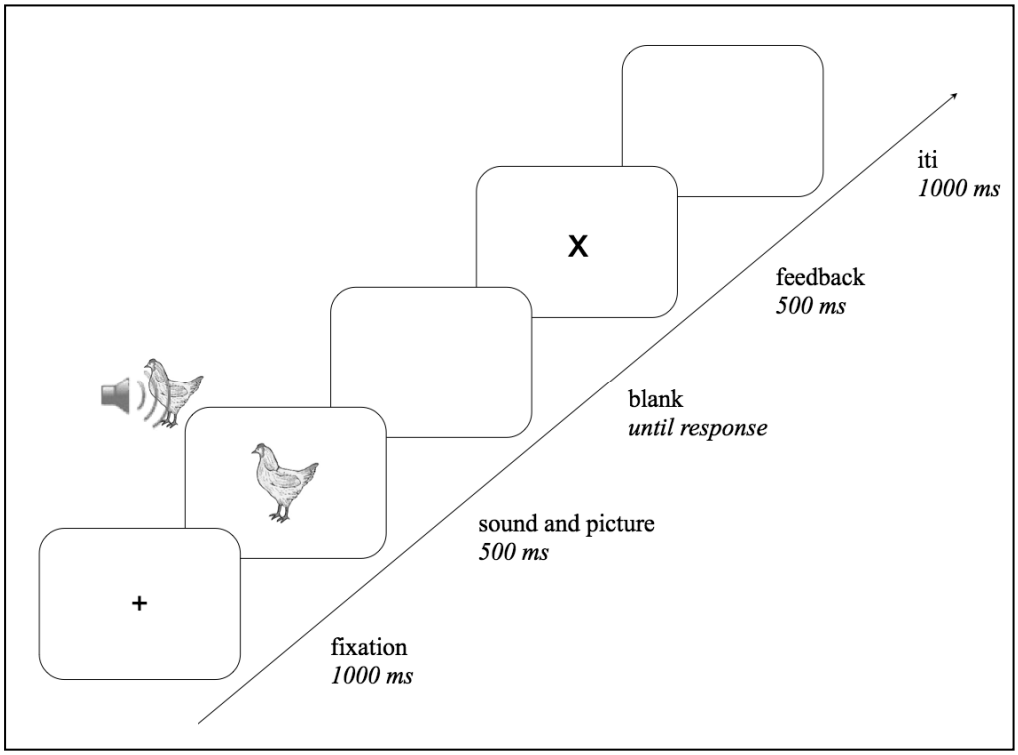


Figure 2.2: Trial events of Experiment 1. After a fixation cross, visual and auditory stimuli were presented (here: stimulus-congruent condition). In case of a incorrect response, feedback was given. In this experiment, a fixed intertrialinterval was implemented.

Data analysis. Trials with stimulus presentation delayed by more than 5 ms were excluded from analysis (less than 1% of all trials) to ensure precise timing. Responses faster than 150 ms or slower than 1500 ms were excluded from further analyses. RTs and error trials were analyzed separately.

For the RT analyses, errors were removed and data were trimmed by 10% at the upper and lower end for each participant, session and condition, which gives more reliable means for the corresponding analyses (Ratcliff, 1993; Wilcox, 1993). Error rates were arcsine transformed before submitting them to statistical analysis. All reported F- and p-values of ANOVAs were corrected after Greenhouse-Geisser.

Error bars in all diagrams represent one estimated standard error of the mean for within-subject-designs (Loftus & Masson, 1994).

Subcategories. Twenty different participants (mean age: 28.7 years) took part in generating subcategories to find semantically near and semantically far conditions. The written German words of all 36 stimuli were printed on paper cards (6.5 x 3.5 cm). Written words were used to ensure a more amodal and comprehensive representation than pictures or sounds. Participants were instructed to generate subcategories separately for living and nonliving categories to receive subcategories for the response-congruent condition. Therefore, each participant was to sort the shuffled word-cards into as many subcategories as he/she pleased. The subcategories were to be chosen subjectively based on semantic content (e.g. not on number of letters). Participants were to provide titles for each subcategory. No time limit was given.

A hierarchical cluster analysis based on the squared Euclidean distance was computed to generate overall subcategories. Results are illustrated in the dendrograms in Figure 2.3a and b. Living items were mostly subdivided into farm animals, wild animals, birds, reptiles and amphibians, as well as insects. Nonliving items were subdivided into musical instruments, objects making skirling noises, vehicles, tools, household items, and a gun. The gun was excluded from further calculations as it was a one-item-category with no semantically near stimuli. A cut-off for generating subcategories was set at a rescaled distance of 11. These subcategories were used to compute semantically near and semantically far conditions within the response-congruent stimuli. Semantically near stimuli consisted of two stimuli from the same subcategory (e.g. sheep and goat), whereas semantically far stimuli originated from two different subcategories (e.g. violin and car).

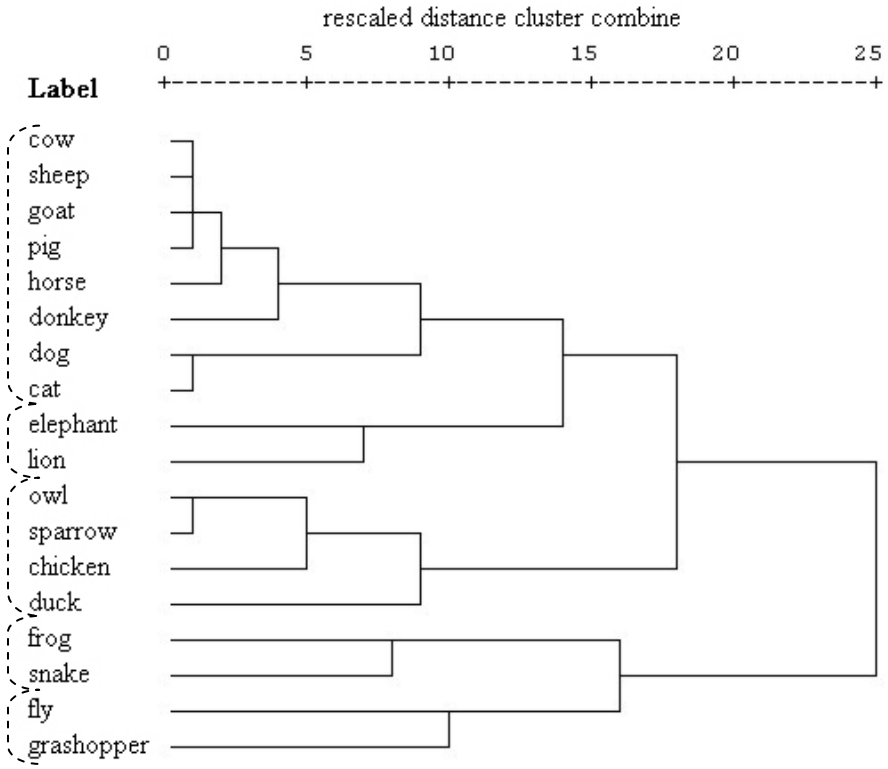


Figure 2.3a: Dendrogram of clusters of words classified as living. Solid lines represent Euclidean distances, dashed lines represent subcategories.

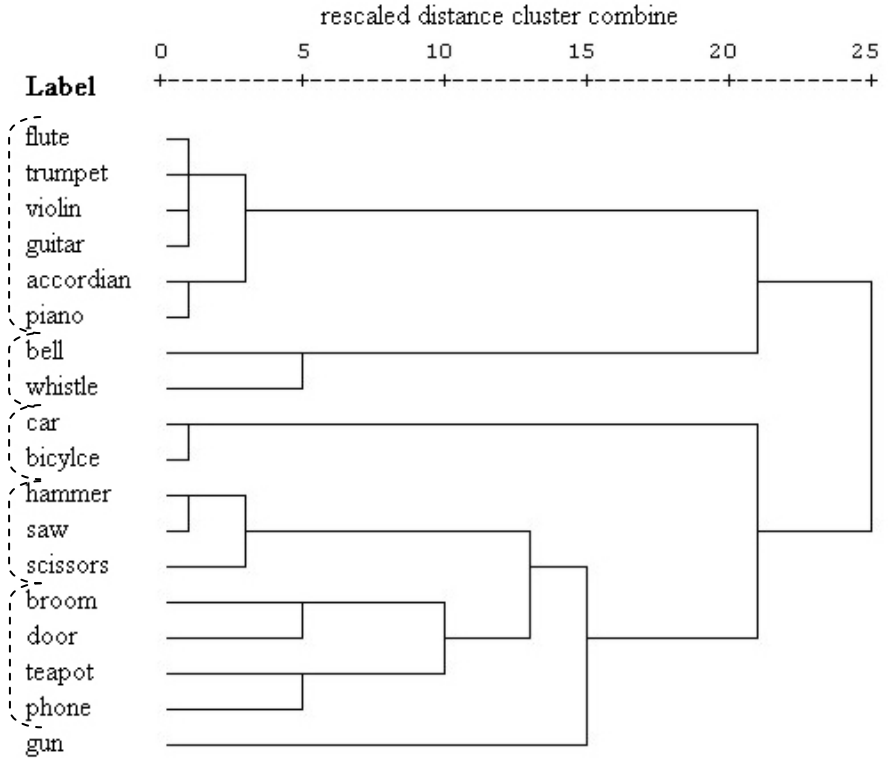


Figure 2.3b: Dendrograms of clusters of words classified as nonliving. Solid lines represent Euclidean distances, dashed lines represent subcategories.

2.1.2 Results

Response times. Participants responded on average 188 ms faster to visual than to auditory stimuli. This significant difference [$t_{(7)} = 5.6$, $p < .005$] is visualized in Figure 2.4. Differences between the two target modalities will not be further analyzed because the hypotheses focus on differences within target modalities. To explore the different influences of the congruence levels, target modalities were analyzed separately. Planned contrasts between the unimodal condition and the other three levels of Congruence Type show if the bimodal stimuli are processed faster or slower than the unimodal stimuli.

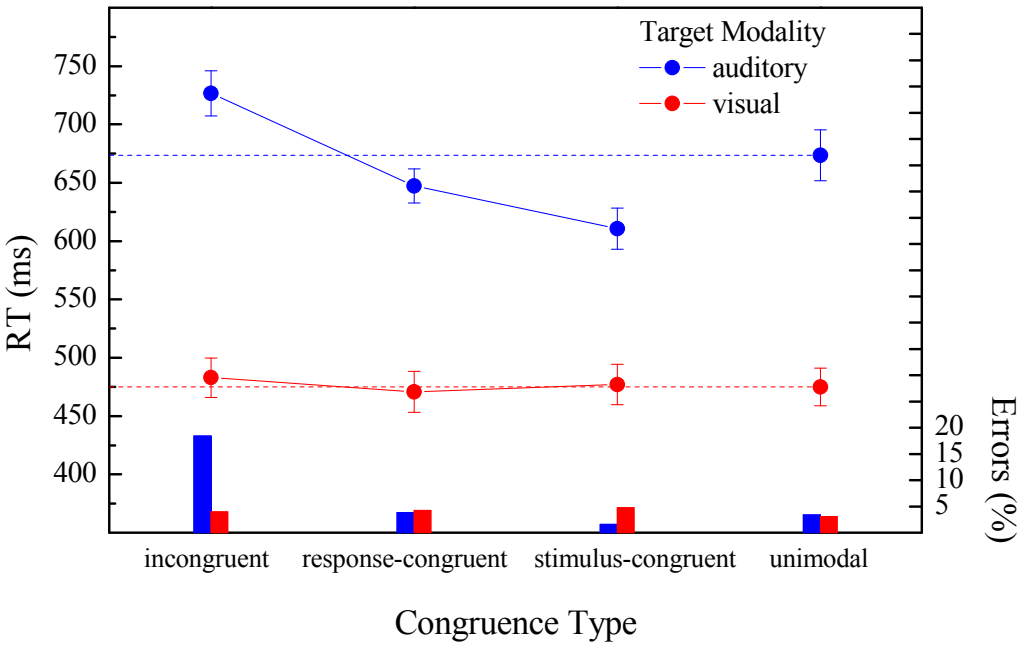


Figure 2.4: Mean RTs (lines) and errors (columns) for auditory and visual targets as a function of Congruence Type. Unimodal conditions served as a baseline and are thus represented by a dashed line. Error bars of RTs represent the standard error of the mean for repeated measures (Loftus & Masson, 1994).

Figure 2.4 clearly shows that when participants attend to the auditory modality, the more congruent the stimuli, the faster the responses. Compared with unimodal auditory stimuli, stimulus-congruence led to faster responses, $F_{(1,7)}=26.4$, $p < .005$. Such a bimodal enhancement was also found for stimuli from congruent categories, $F_{(1,7)}=6.6$, $p < .05$. For incongruent stimuli, responses were slower than to auditory stimuli alone, $F_{(1,7)}=23.0$, $p < .005$. The

difference between stimulus-congruence and response-congruence was analyzed in more detail by a post-hoc t-test, which supported that RTs in the stimulus-congruent condition were faster than those in the response-congruent condition, $t_{(7)}$=-4.3, p< .005.

Figure 2.4 shows no difference in RTs, when Target Modality was visual. This was confirmed by non-significant contrasts of comparison between the unimodal visual stimuli and the three other levels of congruence, namely congruent stimuli [$F_{(1,7)}$= .1, p= .74], congruent category [$F_{(1,7)}$= 0.9, p= .38], and incongruent category [$F_{(1,7)}$= 2.6, p= .15].

Errors. 6.8% and 4.0% errors were made for auditory and visual targets respectively. The arcsine-transformed errors of the unimodal condition were compared to each of the remaining three levels of Congruence Type by planed planned contrasts.

When participants responded to the auditory modality, error rates decreased from incongruent, over response-congruent to stimulus-congruent conditions, which is evident in Figure 2.4 (columns). Participants made less errors to stimulus-congruent stimuli than to unimodal stimuli, $F_{(1,7)}$= 15.6, p< .01. Response-congruent stimuli did not differ from unimodal stimuli, $F_{(1,7)}$= 2.9, p= .13. Incongruent stimuli led to increased error rates compared to unimodal stimuli, $F_{(1,7)}$= 13.9, p< .01. Thus, results of RT analyses were supported. No speed-accuracy trade-off was found.

In contrast to RTs, there were significant differences in error rates for visual targets. Increased error rates compared to unimodal stimuli were found in the stimulus-congruent condition, $F_{(1,7)}$= 5.9, p< .05. The contrast between response-congruence and unimodal conditions was also significant [$F_{(1,7)}$= 8.8, p< .05], indicating more errors for response-congruent stimuli. Errors to incongruent stimuli did not differ from unimodal stimuli, $F_{(1,7)}$= 0.2, p= .71.

Subcategories. Differences between semantically far and near stimuli within response-congruent stimuli (i.e. different or same subcategories) were analyzed by a one-sided t-test. It was expected that responses to semantically near stimuli were faster than to semantically far stimuli. A significant main effect of Semantic Distance supports this hypothesis when Target Modality was auditory,

$t_{(7)}$=2.1, p<.05 , but not when Target Modality was visual, $t_{(7)}$=1.3, p=.12. These differences are illustrated in Figure 2.5. Thus, although stimuli were not stimulus-congruent, semantic content had a small effect on response times (at least visual stimuli influenced auditory perception).

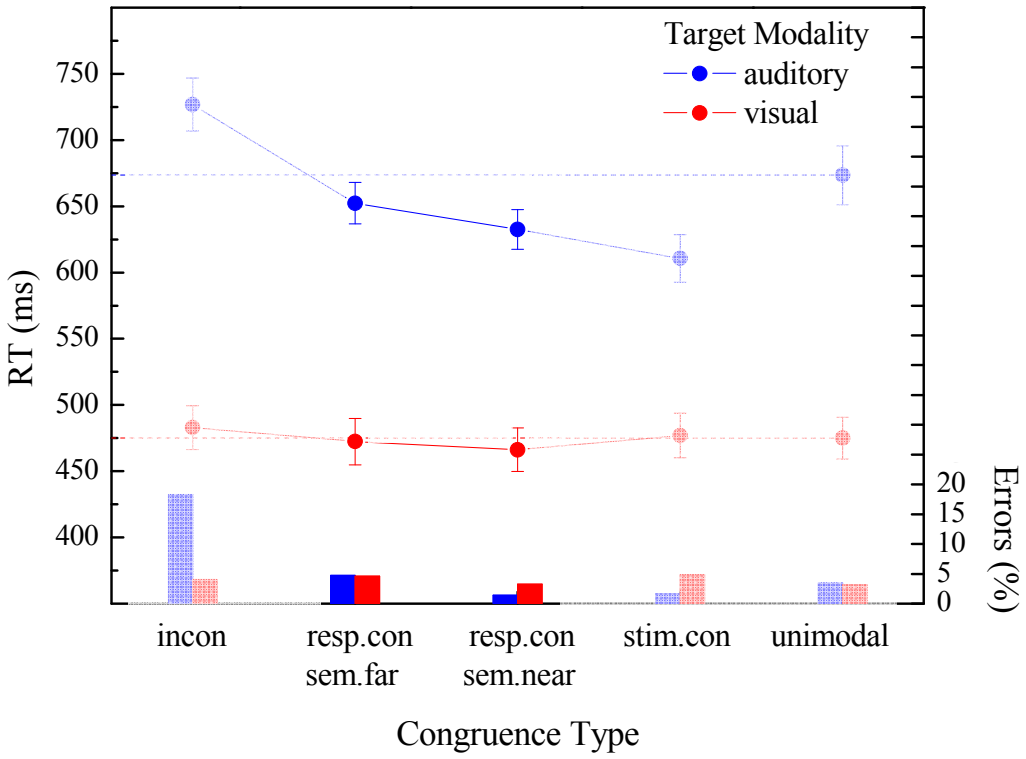


Figure 2.5: RTs (lines) and errors (columns) for auditory and visual targets as a function of Congruence Type. Response-congruence (resp.con) was divided into two conditions, that is semantically near (sem.near) and semantically far (sem.far). Remaining conditions (incon=incongruent; stim.con=stimulus-congruent) are the same as in Figure 2.4 (lightened color). Unimodal conditions served as a baseline and is thus represented by a dashed line.

2.1.3 Discussion

Effects of semantic congruence were found when participants attended to the auditory modality. Facilitation cannot be explained by response-congruence alone, because responses in the stimulus-congruent condition were faster than in the response-congruent condition. Thus, semantics had an effect on stimulus processing. This is especially interesting because stimuli came from different sensory modalities, with different physical characteristics. In psycholinguistic experiments the similarity of auditory and visual stimuli is much greater because orthography and phonology of written and spoken words are highly concordant (Holcomb et al., 2005). Here, stimuli only overlapped in meaning.

Responses to visual stimuli were faster than to auditory stimuli. Semantic effects were only evident for auditory targets. There are two explanations why responses to visual stimuli are faster than to auditory stimuli. First, auditory stimuli might be processed more slowly by the auditory system. However, Woodworth and Schlosberg (1954) found faster detection of auditory than of visual stimuli, which might contradict the present results. Another possibility is that visual stimuli might be detected faster because auditory stimuli are dynamic rather than static. Accordingly, static stimuli could be identified right after the onset, whereas identification of auditory stimuli takes some time. This possibility is further explored in Experiment 5.

Why were there no effects from audition onto vision? A floor effect is one possible explanation for the absence of semantic effects from audition to vision. Identification of visual stimuli might be so efficient that there is no room for further improvement.

Another possible explanation is that visual stimuli are less ambiguous and easier to identify than auditory stimuli. This is supported by less correct identifications for auditory than for visual stimuli in a preexperiment (about 60% vs. 80%). Furthermore, responses to unimodal visual stimuli were on average 200 ms faster than responses to unimodal auditory stimuli. Thus, information of visual stimuli seems to be more reliable for the task than that of auditory stimuli, because of better and easier identification of visual stimuli. For example,

when one hears a sound that is not identifiable but the simultaneously presented picture is clear, the response will be dominated by the unambiguous picture. Correspondingly, Welch and Warren (1980) have found different reliabilities for vision and audition in spatial and temporal tasks. Recently, Banks (2004) argued that the best combination of information from two modalities is a weighted average of both. He found best results from a neural network when responses are based more on highly reliable stimuli than on less reliable ones. A response with such a combination is better than a response to either stimulus alone.

Another possibility for absent effects from audition onto vision is that the stimuli are not processed simultaneously by the brain. Although usually auditory stimuli are processed faster than visual stimuli, visual stimuli were static and auditory stimuli were dynamic. Visual stimuli can be fully identified with their onset, while auditory stimuli need to be presented for some time before it can be identified. Thus, different identification speeds might also reflect differently perceived simultaneities. Effects from vision onto audition were evident because visual stimuli subjectively precede auditory stimuli and therefore prime the target. When exploring effects from audition onto vision, the target is perceived before the prime and no effect is evident. These possibilities are further explored in the following experiments.

Semantic distance had a small effect on judgments when participants attended to auditory stimuli. Responses to semantically near stimuli were faster than to semantically far stimuli. Thus, the brain processes information from two different modalities faster when this information has similar meanings. The hypothesis of influences of semantics on multisensory integration is thereby further supported.

On the other hand, the subgroups have methodical limits because of the post-hoc implementation. Frequencies of stimulus combinations were therefore not balanced. As a result, effects were more vulnerable to error probability. Results of subcategories are to be considered with caution until an experiment with balanced frequencies is conducted.

How can the effect of semantic distance be explained? The spreading activation model of Quillian (1962) gives an explanation on how semantic informa-

tion might be organized in the brain. It assumed that concepts (e.g. meaning of words and items) are represented by nodes in a network (i.e. neurons in the human brain). Semantically near items are represented by stronger connections between these nodes than semantically far items. Thus, faster responses to semantically near stimuli may arise from stronger connections between their nodes.

Follow-up experiments that try to explore effects of semantic distance should introduce a corresponding factor in the experimental design. An appropriate control would be the usage of numeric stimuli. Numbers have the advantage that distance is clearly defined. Reynvoet, Brysbaert and Fias (2002) explored semantic priming for numbers, using digits and word numerals as primes and targets. Their results indicate that the smaller numerical distance between prime and target, the faster the naming of the target number. Thus, as here, distance affected responses. To my knowledge, this design has not yet been used in a crossmodal design.

2.2 Experiment 2: Effect of SOA

Experiment 1 showed that semantic congruence of audiovisual stimuli affect responses. Responses to auditory stimuli were facilitated. Three possible reasons for the absence of influences from audition onto vision were discussed in Experiment 1. First, visual stimuli were so efficient that no further improvement was achieved by congruent auditory stimuli. Second, the visual stimuli were more reliable than the auditory stimuli. Therefore, less reliable auditory stimuli did not affect perception of visual stimuli. Third, auditory stimuli arrived later in the integration sites. Due to possible violation of the temporal rule, no integration occurred.

To explore reasons for the absence of effects, two changes will be implemented in Experiment 2. First, visual stimuli are blurred, such that identification is hindered. A pilot-experiment showed that a Gaussian blur filter of eight pixels led to harder but still sufficient identification. Figure 2.6 shows example stimuli from Experiment 1 and 2. Responses should be more difficult and floor effects should decrease. Additionally, these visual stimuli should evoke less reliable responses. Thus, audition may have an effect onto vision. On the other hand, participants should still be able to identify the stimuli. Therefore, effects from vision to audition should remain as in Experiment 1.

Figure 2.6: Examples of stimuli from Experiment 1 (left) and Experiment 2 (right).

Second, different SOAs will be introduced. SOA varies from 0 ms to 500 ms. For example, a SOA of 200 ms means that the visual stimulus starts 200 ms before the auditory stimuli for auditory targets (vice versa, when Target Modality is visual). If auditory stimuli arrive later in the integration sites, then effects from vision onto audition should be found at larger SOAs. Let's hypothesize, a visual stimulus takes 50 ms to be correctly perceived, while the perception of an auditory stimulus is evolved in 250 ms (e.g. due to a later p-center; Morton,

Marcus & Frankish, 1976). Accordingly, if the auditory stimulus is presented 200 ms before the visual stimulus, they should be perceived as simultaneous and thus be integrated by multisensory neurons. Figure 2.7 illustrates this effect of different p-centers.

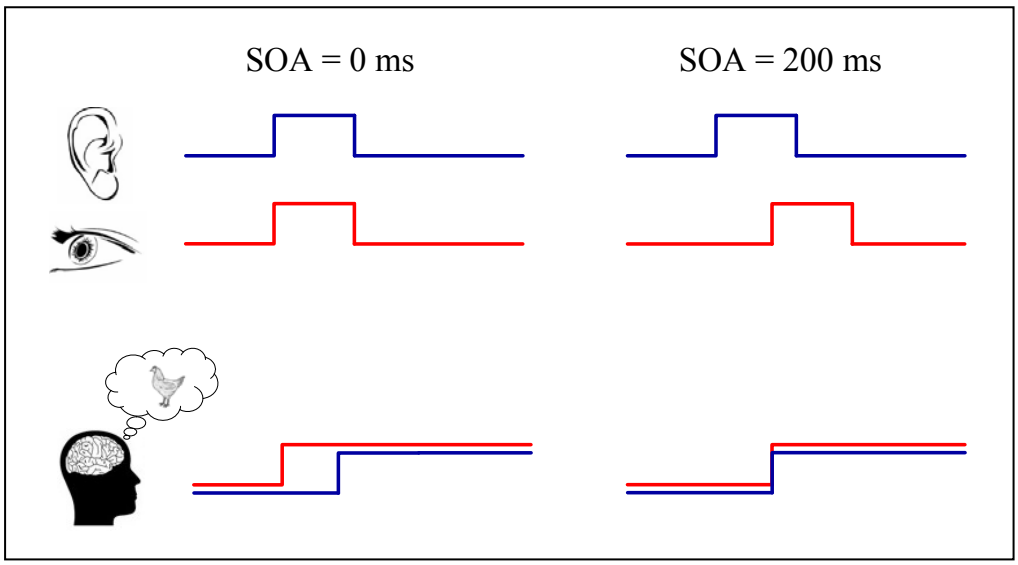


Figure 2.7: Hypothetical demonstration for the perception of auditory (blue lines) and visual (red lines) stimuli. The upper panel illustrates presentation times of an auditory and a visual stimulus for a SOA of 0 ms and 200 ms, while the lower panel illustrates the identification process. Presenting stimuli simultaneously may lead to asynchronous perception (i.e. audition after vision). A SOA of 200 ms may induce simultaneous perception.

By implementing SOAs, it can further be explored whether the processes are automatic or strategic. According to Perea and Rosa (2002), at small SOAs (less than 250 ms), automatic processing dominates, whereas strategic effects dominate at larger SOAs (more than 400 ms). Furthermore, the magnitude of semantic effects increases with SOA (Perea & Rosa, 2002). It is probable that these effects were not caused by strategies alone, because in Experiment 1, a SOA of 0 ms was used and effects of semantic relation were evident. On the other hand, it is also possible that strategic effects play at least a subordinate role, because the stimuli were not masked and are thus processed consciously.

The experimental setup is as in Experiment 1, except for the usage of blurred visual stimuli and SOAs. SOAs ranged from 0 to 500 ms. Stimuli from the target modality are always presented last. It is hypothesized that larger effects of

the unattended modality will be found at larger SOAs. Furthermore, blurred stimuli should result in larger influences of audition onto vision also at an SOA of 0 ms.

2.2.1 Methods

Participants. Eight undergraduate students (one male) received course credit for participation. The average age was 22.3 years (range from 20 to 26). All were naive as to the purposes and hypotheses that motivated the study. They reported normal or corrected-to-normal vision, as well as normal hearing. All participants were right-handed.

Apparatus and Stimuli. All were the same as in Experiment 1, except that visual targets were blurred. The visual stimuli from Experiment 1 were edited with Corel Photo-Paint, Version 10 (cf. Figure 2.6). A Gaussian blur filter was applied to each stimulus. A Gaussian blur filter of 8 pixels led to slower identification in a pilot-experiment. Thus, identification of visual stimuli was more difficult.

Design. Three factors were included in the experiment. As in Experiment 1, Target Modality was visual and auditory. Congruence Type consisted of four levels, that is, stimulus-congruence, response-congruence, incongruence, and unimodal presentation (a single visual or auditory stimulus). Furthermore, SOA was varied with the levels 0, 50, 100, 200, 350, and 500 ms. SOA could not be varied in the unimodal condition. The order of Target Modality was balanced between sessions over participants, whereas the other factors were randomized blockwise. RTs and errors were taken as dependent variables.

Task. The task was the same as in Experiment 1.

Procedure. The only procedural difference to Experiment 1 was the introduction of SOAs. Thus, visual and auditory stimuli were presented simultaneously at a SOA of 0 ms. In the other SOA conditions the stimuli were presented after

one another (SOA = 500 ms) or overlapped in time (50 to 350 ms SOAs). In the latter cases, the target stimulus was always presented last.

Data analysis. RT and error data were treated as in Experiment 1.

2.2.2 Results

The results are summarized in Figure 2.8. Again, Target Modalities were analyzed separately.

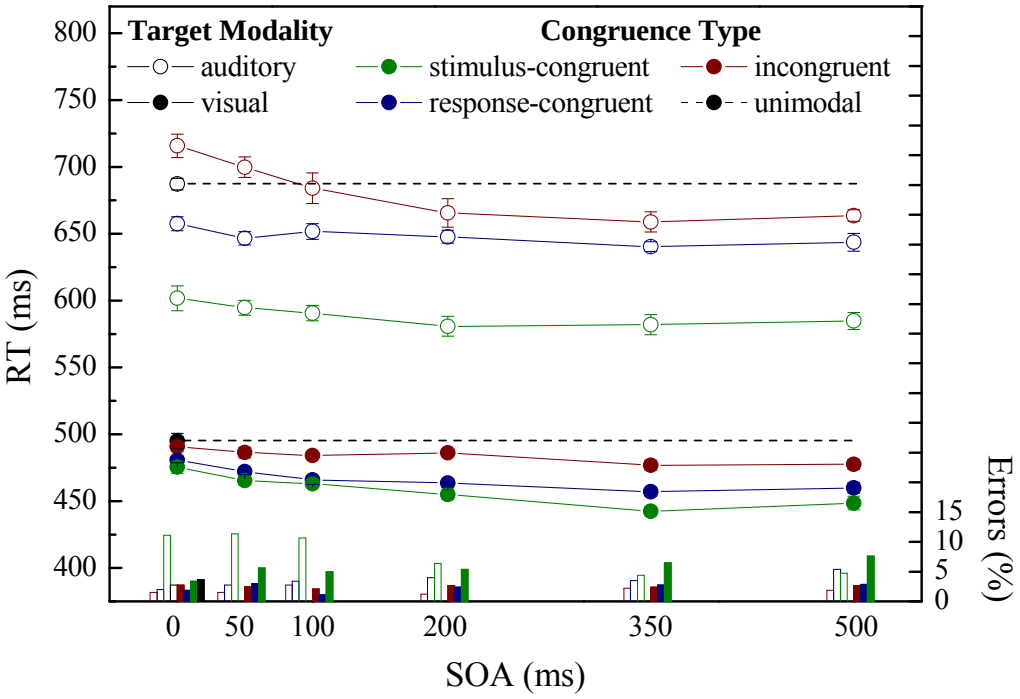


Figure 2.8: RTs (lines) and errors (columns) of Target Modality and Congruence Type as a function of SOA. SOA could not be varied in the unimodal condition (dashed lines), because just one stimulus was present.

Response times. For responses to auditory targets in Figure 2.8, it is obvious that stimulus-congruence elicited the fastest responses throughout all SOA conditions, followed by response-congruent and incongruent stimuli respectively. The differences are supported by a significant main effect of Congruence Type, $F_{(2,14)}= 55.4$, $p< .001$. With increasing SOA, mean RT decreased,

which is evident in a significant main effect of SOA, $F_{(5,35)}$=16.1, p<.001. Nevertheless the differences between the response-congruent condition and the incongruent condition is less distinct with increasing SOA, which is the interpretation of the significant interaction of SOA and Congruence Type, $F_{(10,70)}$=2.9, p<.05.

When participants had to respond to the visual stimulus, effects were smaller than the auditory Target Modality. However, over all SOAs, the stimulus-congruent condition again elicited the fastest responses, followed by response-congruent and incongruent conditions. This is supported by the main effect of Congruence Type, $F_{(2,14)}$=49.6, p< .001. Also similar to the auditory Target Modality, RTs decreased with increasing SOA, $F_{(5,35)}$=13.6, p< .001. Furthermore, the interaction of Congruence Type and SOA was nonsignificant, $F_{(10,70)}$=2.0, p= .12.

Next it was explored whether the blurred visual stimuli had an effect. In order to compare the results to Experiment 1 that included clear pictures, congruence effects in the 0 ms SOA condition was further analyzed by contrasting the three congruence levels with the unimodal condition. Effects attenuated with decreasing congruence. RTs to stimulus-congruent stimuli differed significantly from the unimodal condition, $F_{(1,7)}$=13.2, p< .01. Responses to congruent stimuli differed only marginally [$F_{(1,7)}$=5.0, p< .07] and to incongruent stimuli did not differ significantly from responses to unimodal stimuli [$F_{(1,7)}$=1.0, p= .34]. Thus, in contrast to Experiment 1, auditory stimuli affected perception of visual stimuli.

The net congruence effects are illustrated in Figure 2.9. Category effects were contrasted to semantic effects for both target modalities. Category effects resulted from subtracting all response-congruent conditions (i.e. response-congruence and stimulus-congruence) from incongruent conditions. Semantic effects were calculated from the difference between response-congruent and stimulus-congruent conditions. Results differed in trend for auditory and visual targets. When participants attended to auditory targets, category effects decreased with increasing SOA, while semantic effects rather increased. On the

other hand, semantic and category effects increased for visual targets with increasing SOA.

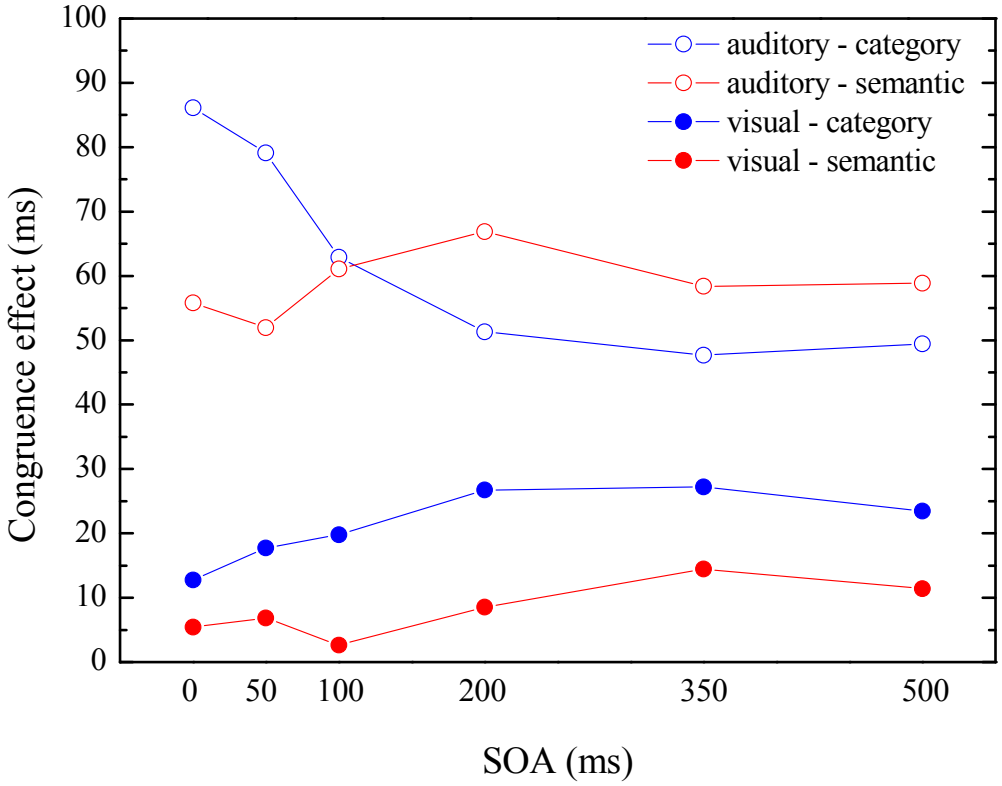


Figure 2.9: Category and semantic effects for visual and auditory targets as a function of SOA. Category effects result from the difference between incongruent and all response-congruent (mean of response-congruent and stimulus-congruent) conditions. Semantic effects are the difference between response-congruent and stimulus-congruent conditions.

Errors. Responding to auditory and visual stimuli, only 4.4% and 3.5% errors on average were made respectively. An ANOVA with arcsine transformed relative frequencies was calculated. The effects are visualized as columns in Figure 2.8. Errors supported RT results. A speed-accuracy trade-off was ruled out, because fewer errors were made when responses were faster.

For auditory targets, error rates declined with levels of Congruence Type, $F_{(2,14)}$=22.6, p< .001. The least errors were made in the stimulus-congruent condition, followed by response-congruent and incongruent conditions. SOA did not significantly influence error rates, $F_{(5,35)}$=1.2, p= .35. The significant interaction [$F_{(10,70)}$= 3.5, p< .02] indicates that errors declined with increasing

SOA in the incongruent condition, however, a slightly increase was evident in response-congruent and stimulus-congruent conditions.

When participants responded to visual stimuli, error rates also declined with increasing congruence, $F_{(2,14)}= 9.2$, $p< .005$. This seems to result mainly from the increased errors in the incongruent condition. Stimulus-congruent and response-congruent conditions were rather similar. The main effect of SOA as well as its interaction with Congruence Type was nonsignificant, $F_{(5,35)}=1.7$, $p= .17$ and $F_{(10,70)}= 1.2$, $p= .35$ respectively.

2.2.3 Discussion

Congruence effects were again larger for auditory than for visual targets. In contrast to Experiment 1, congruence also had an influence on visual perception. Results of this crossmodal priming task showed furthermore that semantic effects rise slightly with increasing SOA, but were in trend evident at all SOAs. Specifically, mean responses on stimulus-congruent trials were always faster than on response-congruent trials, which again were faster than on incongruent trials.

When focusing on the 0 ms SOA conditions, which had the same time interval as in Experiment 1, one can see that congruence effects from audition onto vision were small but pointed in the hypothesized direction. Mean RTs of responses to visual targets were similar to those in Experiment 1, showing that blurring of visual stimuli had a small effect. Thus, reliability changed slightly and floor effects could again not be prevented. Auditory stimuli were harder to identify in the current experiments than visual stimuli (cf. preexperiment in 2.1). As a result, smaller influences from audition onto vision were found.

How can the increase of semantic effects with increasing SOA be explained? One possibility is that at larger SOAs, perceived simultaneity of vision and audition may be achieved. Perceived simultaneity was tested with a temporal order judgment task in Experiment 4. Based on the present experiment, simultaneity seems not to have a large influence, because increasing effects with increasing SOA were evident for both target modalities. Effects were larger

when the distractor preceded the target. Second, strategic effects might be responsible for increased semantic effects at larger SOA (cf. Perea & Rosa, 2002). Which is the more appropriate explanation of the two possibilities? Strategic effects probably had some influence but it is implausible that they produced all effects. Semantic influences increased until a SOA of 350 ms. At a SOA of 350 ms semantic influences slightly decreased again. This decrease does not support strategic effects. Furthermore, semantic influences were at least trendwise evident at all SOAs. Thus, some automatic processing must have occurred. Response priming may explain why semantic effects increase together with SOA. But it cannot be accounted for faster responses over all conditions (including incongruence). A simple priming of any reaction may have led to facilitation of all responses.

2.3 Experiment 3: Detection task

Effects of semantic congruence were found in the previous experiments. These effects could not be explained solely by response-congruence. However, congruence was task relevant, because the unattended modality may have helped or hindered to respond to the target modality. The objective of the present experiment is to find out whether semantic influences will also be observed when congruence is completely irrelevant. Up to this experiment, the neuronal level of integration remained unsolved. Does multisensory integration occur at a low level, as in experiments with simple flashes and clicks? Or can enhancement of responses be explained by a response bias on a higher level? There are two possibilities to solve these questions. First, fMRI or PET studies could reveal which sites are activated during the previously tasks. And ERPs could indicate at what time after stimulus presentation differences between congruence levels arise. Second, congruence effects in a task involving less deep processing would imply that deeper processing is not sufficient for semantic influences. Thus, integration occurs at a rather low level. Not only because of missing technical devices, the latter approach was chosen. But a low level task also implies that stimuli are task-irrelevant, which allows to explore if the semantics of stimuli needs to be relevant for the task. And I can find out if effects of semantic relation are evident in different tasks. One task involving low processing only of the stimuli is a simple detection task (Perea, Rosa & Gómez, 2002).

In the present experiment, participants are instructed to respond as fast as possible when a visual and an auditory stimulus are jointly presented. When either a visual or an auditory stimulus is present, no response is to be given. The semantics of stimuli is thus irrelevant for the task. As no choice is required, less processing is required. If semantics operates on a lower level, differences in RT according to congruence of the stimuli are expected. On the other hand, no effects will be evident if deeper processing is necessary for effects of semantic relation.

2.3.1 Methods

Participants. Eight female undergraduate students received course credit for participation. The average age was 25.8 years (range from 19 to 40). All were naive as to the purposes and hypotheses that motivated the study. They reported normal or corrected-to-normal vision, as well as normal hearing. All participants were right-handed.

Apparatus and stimuli. The equipment and the stimuli remained unchanged from Experiment 1.

Design. All factors were repeated measures. Congruence Type consisted of the levels stimulus-congruent, response-congruent, incongruent, and unimodal. SOA was varied for the bimodal Congruence Type conditions. The SOA levels were -100, -50, 0, +50, +100 ms. Negative values indicate presentation of auditory stimuli before visual stimuli. Vice versa is the case for positive values. RTs served as the dependent variable. A smaller SOA range than in Experiment 2 was used for economy and because the present experiment does not focused on perceived simultaneity.

Task. Participants were instructed to press one of the control-keys as fast as possible when a sound and a picture were presented. If either sound or picture alone was presented, participants were not to respond (no-go) and to wait for the next trial.

Procedure. Every participant attended two 1-hour-sessions, mostly on consecutive days. Each session of 740 trials was split into three blocks, separated by 30-seconds breaks. Each block consisted of 240 trials. Additionally, 20 trials served as practice in the beginning of each session. The practice trials were excluded from any analyses.

A trial was set up as in Experiment 3, except that participants had only 1500 ms to respond. Otherwise, the next trial started automatically. The measuring of RTs started with the onset of the later stimulus.

Data analysis. All error trials were excluded from analyses. Due to the go/no-go design, unimodal trials could not be analyzed. RT data was treated as in Experiment 1.

2.3.2 Results

A repeated-measure ANOVA included SOA and Congruence Type. Figure 2.10 shows that the smaller the temporal difference between the two stimuli, the slower were the responses, which was evident in a significant effect of SOA, $F_{(4,28)}=8.8$, $p< .005$. In other words, RTs decreased the more one of the stimuli was presented before the later one. SOAs of ± 50 ms led to shorter RTs. Fastest responses were found at SOAs of ± 100 ms. Congruence Type did not influence RTs significantly [$F_{(2,14)}=1.9$, $p = .19$], which is observable in Figure 2.10. Furthermore, the interaction of SOA and Congruence Type was not significant, $F_{(8,56)}=1.2$, $p = .34$.

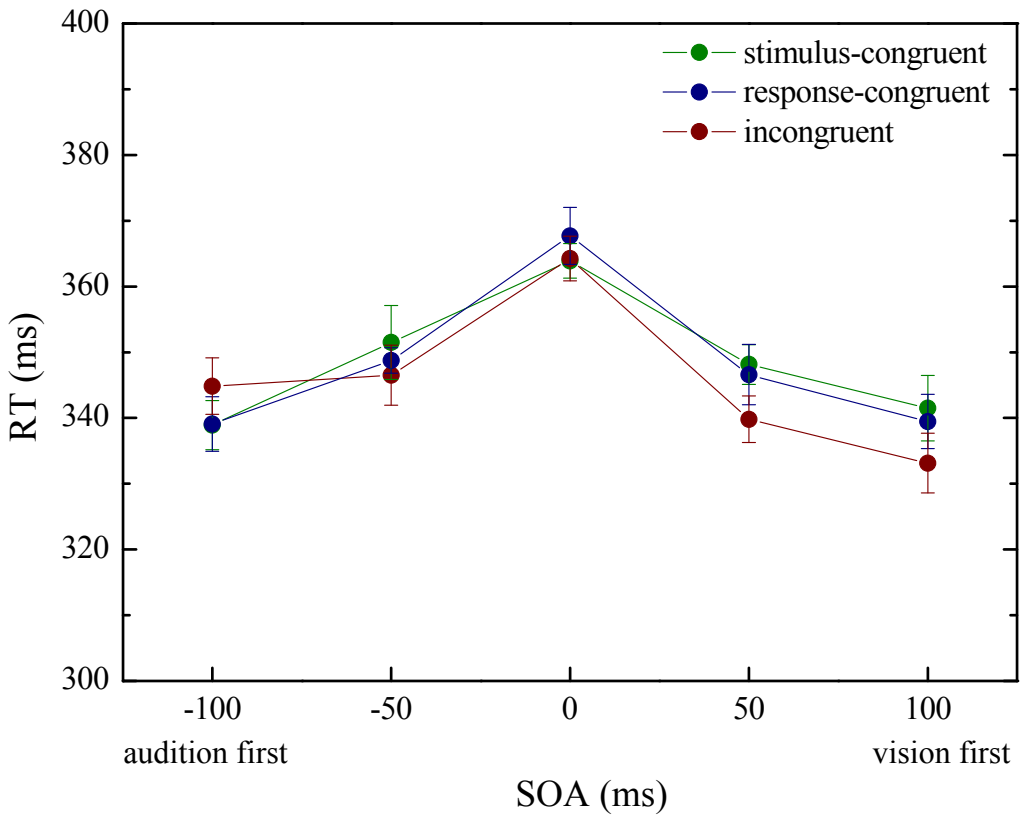


Figure 2.10: RTs of Congruence Type levels as a function of SOA in Experiment 3.

2.3.3 Discussion

No effect of semantic congruence was found with this detection task. Thus, processing of semantics requires deeper processing. Here, the detection task required lower processing (Perea, Rosa & Gómez, 2002). Does this further imply that task irrelevant stimuli have no semantic effect? This remains an open question, because the effects of task relevance and processing level cannot be separated. Neuroimaging studies would help to distinguish these effects.
The effect of SOA might be explained by response priming. Responses are usually facilitated when two response-congruent stimuli are presented successively. The longer the interval between the stimuli, the larger are response priming effects (e.g Vorberg, Mattler, Heinecke, Schmidt & Schwarzbach, 2003, for unimodal response priming). In the present experiment just response on detection was requested. Thus, response priming was also evident for incongruent stimuli.
A smaller range of SOAs was used here, as compared to Experiment 2. However, larger semantic influences were found in Experiment 2 for larger SOAs. Thus, semantic influences might have been found, if a larger SOA range had been used. Experiment 4 tests this hypothesis by implementing a larger range of SOAs. Furthermore, a different low-level task is used.

Because the absence of semantic effects resembles the null hypothesis, a further experiment with a single reaction time paradigm was conducted. The task was simply to respond as fast as possible to any stimulus. It is not included here, because it revealed no further information and had less statistical power due to fewer trials. The only new result was that responses to unimodal stimuli (a visual or an auditory stimulus alone) were generally slower than to multisensory stimuli. This deceleration of responses to unimodal stimuli was consistent with former experiments of multisensory integration using more simple stimuli (e.g. Stein & Meredith, 1993). An additional explanation is that priming might not facilitate responses when a single stimulus was presented.

2.4 Experiment 4: Temporal order judgment

In Experiment 2, effects of semantic congruence at several different SOA levels were observed. But are visual and auditory stimuli actually perceived as simultaneous when they are physically simultaneous? Up to now this remains an open question. According to the temporal rule, the neuronal system integrates multisensory stimuli only if they are temporally close (Stein & Meredith, 1993). Otherwise, they are not perceived as originating from the same object. But at which interval were the present stimuli actually perceived as simultaneous? This brings the role of the p-center back into account (cf. Experiment 2). In contrast to visual stimuli, auditory stimuli are dynamic and one might argue that the process until the stimulus is perceived might take longer. On the other hand, auditory stimuli are in general processed faster by the neuronal system than visual stimuli, as indicated by lower RTs (Woodworth & Schlosberg, 1954).

The previous experiments leave further open questions. One question is at which processing stage multisensory signals are integrated. This question was discussed in 1.2.2 and the answer was mainly mix between early and late processing. Is this also the case for multisensory effects of semantic congruence? One way to solve this is to use a task that requires less intensive processing. If effects of semantic congruence are found in a task that does not require semantic processing, this is strong evidence for integration at a relatively early processing stage. This hypothesis was also explored in Experiment 3.

These two approaches can be integrated in an experiment using temporal order judgments (TOJ). In a typical TOJ task, the SOA between two successive stimuli is varied, and participants indicate which stimulus was presented first (Keetels & Vroomen, 2005). Using audiovisual stimuli, this means whether the auditory or the visual stimulus was presented first. Other studies directly asked for simultaneity judgments (e.g. Fujisaki et al., 2004). The point of subjective simultaneity (PSS) and the just noticeable difference (JND) are the main statistical measurements in TOJ tasks. The PSS indicates the SOA that corresponds to equally probable responses (i.e. 50% of responses ‘stimulus A was first’ and 50% ‘stimulus B was first’). The JND is defined as the SOA, at which partici-

pants correctly judge the order in 75% of trials (Zampini, Shore & Spence, 2003).

The results of TOJ studies with audiovisual stimuli differ in the direction of their mean PSS. Zampini, Guest, Shore and Spence (2005) found that the auditory stimulus needs to be presented slightly before the visual stimulus, in order to perceive the two as simultaneous. Jaśkowski, Jaroszyk and Hojan-Jezierska (1990) presented the auditory stimulus after the visual stimulus for simultaneous perception. Arrighi, Alais and Burr (2006) used real-world stimuli (i.e. videos and sounds of conga drummers) and also discovered a delay of the auditory stream in order to perceive sight and sound as simultaneous.

Congruence of spatial positions affects simultaneity judgments of audiovisual stimuli (Zampini et al., 2003). Responses are less accurate (i.e. smaller JND) when audiovisual stimuli are presented at the same location than when locations differs. Furthermore, accuracy increases with increasing spatial disparity and especially when hemifields are crossed (Keetels & Vroomen, 2005). Other stimulus properties such as intensity (Roufs, 1974), frequencies and color (Fink, Ulbrich, Churan & Wittmann, 2005) as well as changes of direction of movement (Arrighi, Alais & Burr, 2004) can also influence TOJs. Top-down strategies also influence simultaneity judgments. Zampini, Shore and Spence (2005) manipulated attentional mechanisms by instructing participants to either attend to vision or to audition. Auditory stimuli needed to precede visual stimuli for both target modalities. But when participants attended to auditory stimuli, the interval between the stimuli needed to be larger for correct TOJs than when the visual modality was attended.

The perceptual system can recalibrate simultaneity. Fujisaki et al. (2004) presented audiovisual stimuli separated by a fixed interval for several trials. Afterwards they presented the stimuli with different SOAs, while a stimulus set from the adaptation phase was presented once again before every test trial. The results showed that participants moved their simultaneity judgments towards the interval in the adaptation phase. A similar recalibration was discovered with environmental stimuli (i.e. sight and sound of a piano player as well as a speaker) (Navarra et al., 2005). What is the purpose of this recalibration? Sugita and Suzuki (2003) stated that the brain needs to calibrate simultaneity judgments, because at larger distances, sound arrives later than vision (due to slow-

er sonic speed than speed of light). They conducted an experiment with light flashes from different distances (1-50 m) and sound bursts via headphones. The results indicate that the estimated distance and thus the sonic speed affects TOJs.

The present experiment uses the ten visual and ten auditory stimuli that were identified correctly most often in a pilot-experiment (cf. 2.1.1). The stimuli are presented at different SOAs, similar to Experiment 2. Participants are to indicate whether the visual or the auditory stimulus was presented first. According to Jaśkowski et al. (1990), the SOA of perceived simultaneity should be positive (i.e. audition after vision), because auditory signals are processed faster by the brain. Reckoning the p-center-theory (Morton et al., 1976), negative SOAs should be perceived as simultaneous. Another hypothesis regards the effects of semantic congruence. When semantic content has an effect on this low-level processing task, PSS should be smaller and JND should be larger for congruent than for incongruent stimuli.

2.4.1 Methods

Participants. Ten undergraduate students (three male) received course credit for participation. The average age was 22.9 years (range from 20 to 33). All were naive as to the purposes and hypotheses that motivated the study. They reported normal or corrected-to-normal vision, as well as normal hearing. Every participant reported right-handedness.

Apparatus. The equipment remained unchanged from Experiment 1.

Stimuli. All stimuli were the same as in Experiment 1, except for the following changes. The present experiment included only the stimuli that were correctly classified and identified most often (i.e. ten stimuli classified as living and ten stimuli classified as nonliving).

Design. All factors were varied within subjects. Congruence Type consisted of the levels incongruent, response-congruence and stimulus-congruence. Another

factor was SOA, with twelve levels (±500, ±350, ±200, ±100, ±50 and ±10 ms). Reported temporal order (picture first vs. sound first) served as the dependent variable.

Task. Participants were to indicate as accurately as possible whether picture or sound was presented first. No time limit was given. Left and right control-keys served as response buttons. The order was balanced over participants. Error feedback was not included.

Procedure. Every participant attended two sessions á 720 trials subdivided in four blocks each. Each combination was thus presented 24 times. Additionally, 40 randomly chosen practice trials were presented in Session 1 and 20 in session 2. Practice trials were excluded from any calculations.
A trial started with presentation of a fixation cross (size: 0.3° x 0.3°) for 500 ms. Afterwards, a visual (size: 8.2° x 8.2°) and an auditory stimulus were presented for 500 ms, whereas the order and interval was varied with SOA. Following a response, the intertrialinterval was randomly chosen between 1000 and 1500 ms. The next trial started automatically.

Data analysis. Data were not trimmed by RTs because no speeded response was required. One participant was excluded from any analyses, because his/her point of subjective simultaneity (PSS) and his/her just noticeable difference (JND) lay outside the implemented SOA levels.

2.4.2 Results

Proportions of 'visual first' responses (i.e. higher scores indicate that the visual stimulus appeared longer before the auditory stimulus) were calculated for each participant and condition. These proportions were transformed to equivalent Z-scores, assuming an inverse cumulative normal distribution. This procedure is analogous to probit analysis (Finney, 1964) and helps to analyze discrete data with linear regressions. Next, linear regressions were performed per participant and Congruence Type level. SOA levels outside ± 200 were excluded, because

performance was nearly perfect and thus no additional variance was included (cf. Zampini, Shore & Spence, 2003). R^2 values of all regressions were above .75. The resulting slopes and intercepts of the best-fitted lines were used to calculate the PSS and JND. The PSS is the value that is equally probable for 'visual first' and 'auditory first' responses. It was computed by dividing the negative slope by the intercept. The JND indicates at which SOA participants were correct on 75% of the trials. Accordingly, 0.675 (which is the Z-score that corresponds to .25 and .75 in relative frequencies) was divided by the slope. ANOVAs were used to explore differences between Congruence Type levels, separately for PSS and JND.

This procedure was also used by several other TOJ tasks using multisensory stimuli (e.g. Zampini et al., 2005; Keetels & Vroomen, 2005; Sinnett, Juncadella, Rafal, Azañón & Soto-Faraco, 2007). Mean responses to the stimuli of all Congruence Type levels at the different SOAs are shown in Figure 2.11. Regression lines are also included.

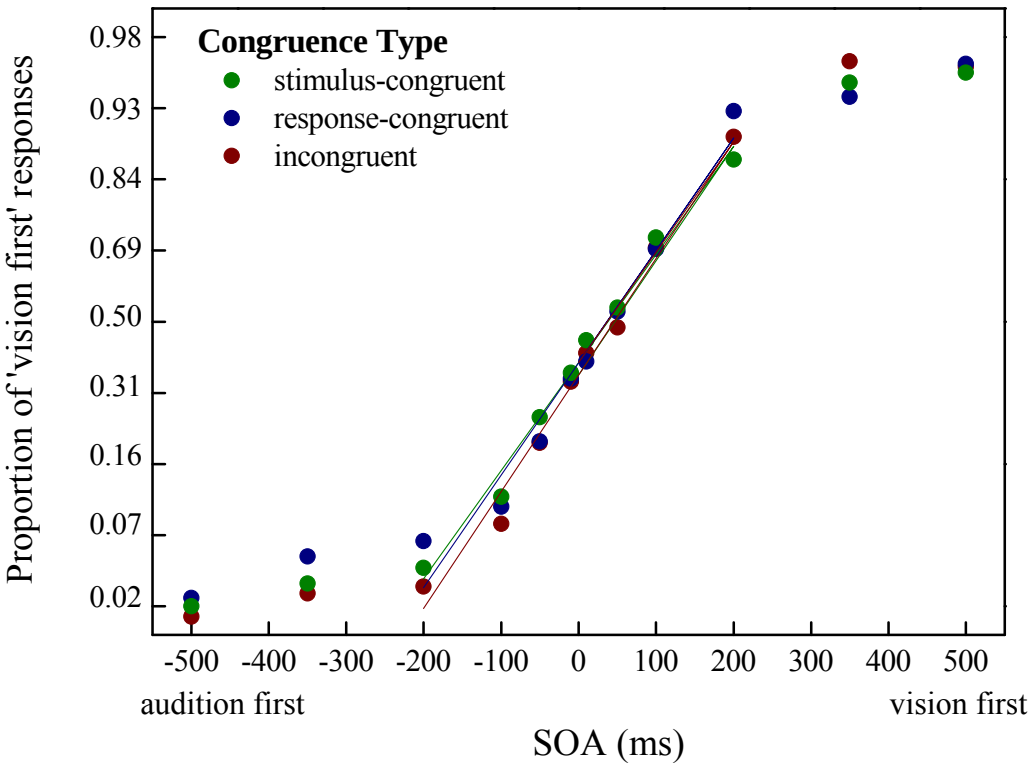


Figure 2.11: Mean proportion of 'vision first' responses plotted on a scale of normal distributed Z-scores as a function of SOA and Congruence Type. Best-fitting regression lines were calculated for intermediate SOAs (± 200 ms) separately for Congruence Type levels.

Mean PSS and JND were listed in table 2.2. Over all conditions, the stimuli were perceived as simultaneous when the visual stimulus precedes the auditory stimulus by approximately 40 ms. Temporal differences were only noticeable when the stimuli were separated by at least 90 ms.

Incongruent stimuli were perceived as simultaneous at a larger gap between audition and vision. However, all differences in PSS failed to reach significance, $F_{(2, 16)}= 0.8$, $p= .45$. JND was largest for stimulus-congruence. Again, none of the differences reached significance, $F_{(2, 16)}= .9$, $p= .4$.

Table 2.2: Points of subjective simultaneity (PSS) and just noticeable differences (JND) as a function of Congruence Type as well as overall means of Experiment 4.

	PSS (ms)	JND (ms)
stimulus-congruent	37.66	97.10
response-congruent	36.21	91.66
incongruent	45.27	91.48
Mean	39.71	93.41

2.4.3 Discussion

Participants perceived the employed stimuli as simultaneous, if the visual stimulus preceded the auditory stimulus by about 40 ms. This is analogous to the hypothesis that processing visual stimuli takes longer than processing auditory stimuli (Woodworth & Schlosberg, 1954). Other TOJ studies support this conclusion (e.g. Jaśkowski et al., 1990). Conversely, Zampini, Guest, Shore and Spence (2005) found that visual stimuli need to be delayed for simultaneous perception. The previous experiments showed that responses to auditory stimuli were more slowly than to visual stimuli. The present experiment could explain that increased RTs to auditory stimuli did not imply slower processing of auditory stimuli. On the other hand, participants did not have to identify the stimuli in order to perform the task. Mean RTs of 833 ms conflict with this hypothesis. This seems enough time to enable identification, especially as responses were even slower than in Experiment 1.

Furthermore, the present experiment illustrated that stimuli in Experiment 1 were presumably also perceived as simultaneous. This statement is supported by the large JND of 93 ms on average. Taking the PSS into account, stimuli were perceived as simultaneous from an SOA of -53 to +132 ms. Thus, physical simultaneity was included in perceptual simultaneity. This simultaneity interval could not explain the increased effects at larger SOA in Experiment 2. It was hypothesized that perceived simultaneity led to the increasing effect. But larger SOAs (>150 ms) were not included in his simultaneity interval.

Another main finding was the absence of differences between Congruence Type levels for PSS and JND. Although the effects of PSS and JND were not significant, the differences in trend made sense. The slightly increased PSS for incongruent stimuli indicated that congruent stimuli were perceived as being temporally close because of their semantic congruence. Furthermore, JND was largest for congruent stimuli. As the interval between a visual and an auditory stimulus may be larger when stimuli were congruent, semantics may have had a small, trend-wise influence in this task. However, this interpretation is far from statistical validation and individual data did not provide a clear picture. To test this hypothesis, further investigations require stimuli with less variance, because responses largely depended on randomly drawn stimulus pairs. Another explanation of the absence of effects of semantic relation is that the current task did not require semantic processing. Stimuli identification was not needed to respond correctly. Imaging studies might help to solve the problem.

3. Semantic congruence of movement of simple stimuli

3.1 Experiment 5: Congruence of pitch and movement direction

The previous experiments have shown effects of semantic congruence between pictures and environmental sounds. Effects were limited to a categorization task but could not be explained by response-congruence. The following experimental series will expand these results and avoid critical issues of the previous experiments. The previous experiments contained dynamic auditory and static visual stimuli. One may argue that static visual stimuli are not compatible with dynamic auditory stimuli. How can the picture of a telephone ring? Thus, multisensory integration may be questioned. The results, on the other hand, showed congruence effects between vision and audition. Static stimuli (pictures) strongly affected the perception of dynamic stimuli (sounds). This incongruence of stimulus properties may also be a critical point of the previous experiments. Furthermore, the employed stimuli were very differently within each modality. For example, the sound of a dog was easier to identify than the sound of a broom. Effects might be kept artificially low due to a large variance of identification rates. Stimulus combinations for the response-congruent condition were generated randomly, and not every combination occurred equally often over participants.

An open question is also the neuronal stage at which integration proceeds. The detection task experiment lacked clear evidence of early integration. Again, large variance of the stimuli may play a role here. Another possible explanation is the complexity of the employed stimuli. One might suggest use of more simple stimuli.

Accordingly, I looked for dynamic visual and auditory stimuli that were relatively simple and whose semantic congruence is variable. A study of Maeda, Kanai and Shimojo (2004) used stimuli which might bring a solution to these problems. They presented an ambiguous drifting grating together with a sound which ascended in pitch, descended or had fixed pitch. For a tone with ascend-

ing pitch, participants mostly saw upward motion of the grating, whereas a descending tone had the opposite effect.

A critical point of the study is that participants reported their perceptions. By using this highly subjective measure, participants could just have reproduced what they heard, because the visual stimulus was ambiguous. The authors implemented linguistic stimuli (Japanese words for 'up' and 'down') in a control experiment, however, in which no influence of the words was found. The absence of effects of linguistic stimuli can also be explained by assuming that the relevance of words was obvious and participants responded according to their expectations. Maeda et al. could not reject this assumption, especially as they used subjective report as the dependent variable.

In the previous experiments, semantic effects were found mainly from vision onto audition. Maeda et al. (2004) found effects in the opposite direction, i.e. from audition onto vision. Kitagawa and Ichihara (2002) expanded these results. They presented a square which moved in depth as visual stimulus, simultaneously with a tone. Results revealed an influence from the movement of the square on the perceived change in loudness of the tone. Thus, effects of simple stimuli from vision onto audition and vice versa were found in different studies. Another objective of the present experiment is to find effects in both directions from the same stimuli.

When modifying the experiment of Maeda et al. (2004) in a few points, the above mentioned criteria may be met. Therefore, the visual stimuli will be changed to disks moving upwards, downwards or remaining stationary. Accordingly, tones with ascending, descending and fixed pitches are employed. This change has the advantage that congruence effects in both directions (i.e. vision onto audition and vice versa) may be investigated. As in the previous experiments, participants are instructed to attend to one modality. Participants' task is to indicate as fast as possible whether the visual (or auditory) stimulus moved upwards, downwards or remained stationary. As previously, target modalities are varied over sessions. Further previously mentioned requirements are met by embedding rather simple stimuli and presenting all stimulus combinations equally often.

I predicted effects of semantic congruence occur from vision onto audition and vice versa. RTs in the congruent conditions should thus be smallest. These re-

sults would help to expand and support the findings of Maeda et al. (2004) and of the previous experiments.

3.1.1 Methods

Participants. Eight undergraduate students (two male) received course credit for participation. The average age was 19.9 years (range from 19 to 22). All were naive as to the purposes and hypotheses that motivated the study. They reported normal or corrected-to-normal vision, as well as normal hearing. Six participants were right-handed.

Design. As in the previous experiments, the order of Target Modality was balanced between sessions and over participants. Thus half of the participants responded to visual stimuli in session 1 and to auditory stimuli in session 2, while the order of the other half of participants was vice versa. Levels of Congruence Type between vision and audition were congruent (both stimuli moving in the same direction), incongruent (one moving upwards, one downwards) and neutral (a stationary stimulus in the distractor modality). Again RTs and errors were dependent variables.

Apparatus. The same soft- and hardware as well as the same room were used as in Experiment 1.

Stimuli. Visual and auditory stimuli were presented with varying Congruence Type. Visual stimuli consisted of a black disk with a diameter of 0.8°. This disk either moved upwards or downwards for 200 ms with a speed of about 5.5°/sec, or remained stationary for the same time. An auditory stimulus was presented simultaneously for 200 ms. Three different sounds were chosen to vary congruence between vision and audition. Thus, the pitch either moved upwards (300 to 2000 Hz), downwards (2000 to 300 Hz) or remained the same (1150 Hz) for 200 ms.

Task. Participants were instructed to respond to visual and to auditory stimuli separately in two sessions. The order was balanced over participants. The task was to indicate by pressing a response button whether the participant observed an upwards movement, a downwards movement and no movement of the stimulus in the target modality. Arrow keys (left, down and right) on a standard computer keyboard were marked and served as response buttons. The middle key (down arrow) represented no movement, while the assignment of the left and right arrow-keys was balanced over participants. Thus, half of the participants used the left key to indicate downwards movement and the right key to indicate upwards movement while the other half used the opposite order. It was pointed out that the stimulus in the distractor modality was irrelevant but was supposed to be perceived. Consequently, participants were to watch the screen and wear the headphones throughout the experiment.

Procedure. In each of the two sessions, seven blocks with 120 trials each were presented. Thus, each stimulus occurred 560 times throughout the experiment. Instructions were presented onscreen, followed by 18 practice trials.
Trials started with a fixation cross (size: 0.7°) for 500 ms. Afterwards a visual and an auditory stimulus were presented simultaneously for 200 ms. A blank screen appeared after the response was given. Errors were signaled by presenting a red X (size: 0.7°) for 500 ms. During the following intertrialinterval a blank screen was shown for 1000 to 1500 ms (randomized length).

Data analysis. RT and error data were treated as in Experiment 1.

3.1.2 Results

Response times. Figure 3.1 shows that for both target modalities, congruent conditions elicited fastest responses. Thus, when auditory and visual stimuli moved upwards, responses were faster compared to an upward movement of the visual stimulus and a downward movement of the auditory stimulus. Target modalities were analyzed separately.

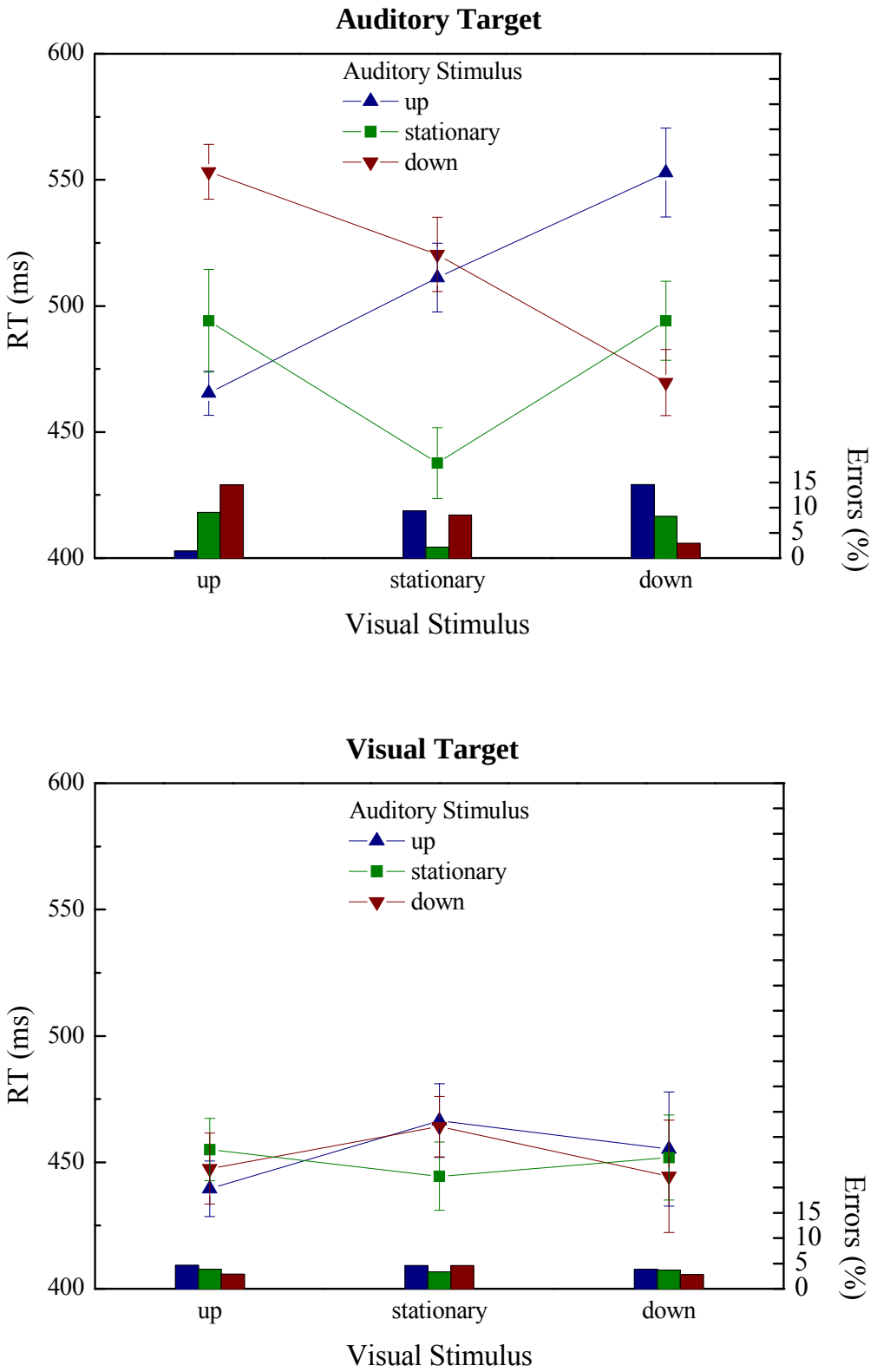


Figure 3.1: Mean RTs (lines) and errors (columns) of Auditory Stimulus as a function of Visual Stimulus. The upper graph contains results for auditory targets and the lower graphs for visual targets.

When participants responded to the auditory modality, the unattended modality had a strong influence on responses. Even though participants were instructed to attend to the auditory stimulus, the visual stimulus influenced the judgments of the auditory stimulus, which resulted in a significant main effect of Visual Stimulus, $F_{(2,14)}=9.6$, $p<.01$. As expected, the auditory stimulus also influenced judgments [main effect of Auditory Stimulus, $F_{(2,14)}=4.6$, $p<.05$]. The significant interaction [$F_{(4,28)}=29.4$, $p<.001$] can be explained by looking at the upper graph of Figure 3.1, where can be seen that the congruent conditions led to the fastest responses. For example, auditory for presentation of an upward movement, an upwards moving disk led to faster responses than a stationary (neutral) disk. Accordingly, incongruent movement (i.e. disk moving downwards) elicited slower responses. Analogously, this trend was also present for a downwards moving sound. For a stationary sound, fastest responses were found with a stationary disk. In this case, no difference is apparent between upwards and downwards movement.

Effects were similar in trend for visual targets. The target stimulus (here: visual) did not influence RTs significantly, $F_{(2,14)}=0.2$, $p=.71$. Neither did the distractor stimulus, $F_{(2,14)}=0.3$, $p=.64$, indicating that overall the auditory stimulus did not influence judgments of the visual stimulus. However, the interaction reached significance, $F_{(4,28)}=6.6$, $p<.01$. This can be best explained by observing the lower graph in Figure 3.1. Congruent stimuli (e.g. visual and auditory upwards movement) led to faster responses than incongruent stimuli (e.g. visual upwards and auditory downwards movement). Effects for responses to visual stimuli were smaller than for responses to auditory stimuli, but effects point in the same direction. For example, the difference between upwards moving auditory stimulus and a downwards moving visual stimulus is 87 ms when responding to auditory stimuli and 16 ms when responding to visual stimuli

Each stimulus combination was presented 90 times per session. Even though every combination was presented equally often, it is possible that participants learned the association between visual and auditory stimuli. If so, effects might have been induced by repeated presentations. To rule out learning effects, congruence effects were plotted for each block separately in the first session (see

Figure 3.2). Incongruent stimuli consisted of visual upwards and auditory downwards motion, or vice versa. Congruent stimuli were defined as congruent motion directions (i.e. up/up and down/down). Combinations with at least one stationary stimulus were excluded, in order to keep frequencies of congruent and incongruent stimuli constant.

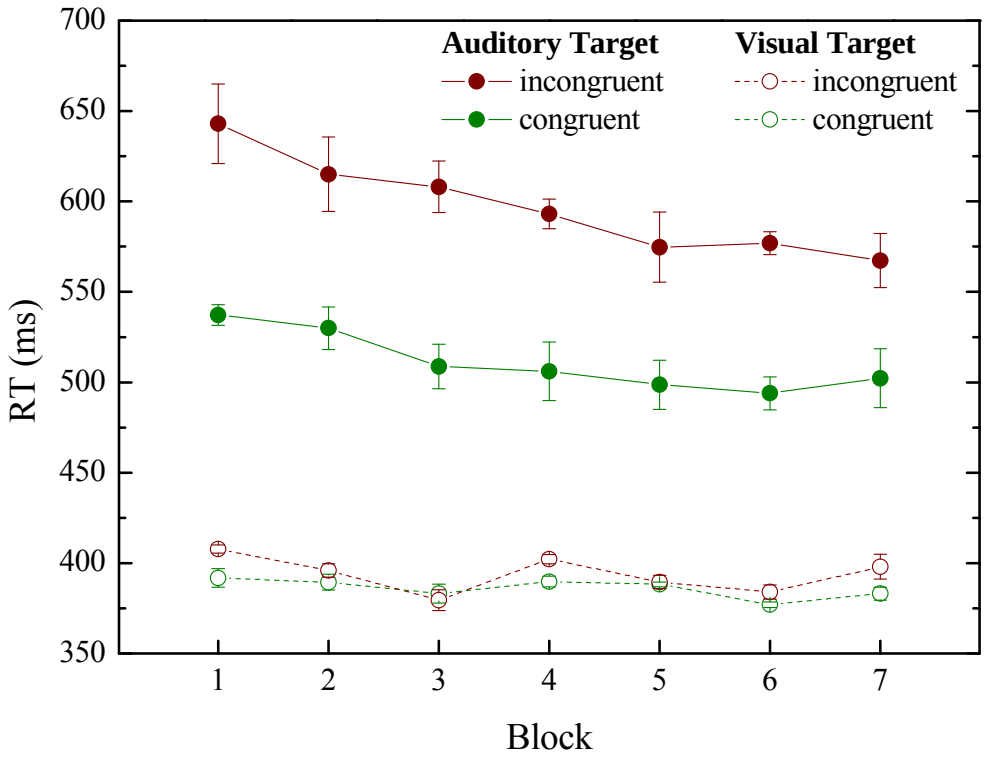


Figure 3.2: Mean RTs for congruent and incongruent conditions of Experiment 5 for separate blocks of session 1. Visual upwards motion plus audition upwards motion as well visual downwards plus auditory downwards motion are congruent stimuli (green lines). Incongruent stimuli (red lines) consist of visual upwards and auditory downwards motion as well as visual downwards and auditory upwards motion. Solid lines represent auditory targets and dashed lines represent visual targets.

For statistical validation, t-tests for repeated measures were computed for the first block separated for target modalities. When Target Modality was auditory, congruence effects were already found in block 1, $t_{(3)}= 4.2$, $p< .05$. For visual targets, effects were again smaller. The difference between congruent and incongruent conditions was only marginally significant, $t_{(3)}= 2.5$, $p< .1$.

Errors. Error data are shown in Figure 3.1 as columns below RT lines. It is obvious that errors run mostly concurrently to RTs and thus indicate similar

underlying effects. No speed-accuracy trade-off was found. On average, 5.8% incorrect responses were made. Thus, differences were rather small. Statistical analysis by an ANOVA for repeated measures resulted in non-significant effects for all factors and interactions (all $p>.3$) except a significant interaction between Visual Stimulus and Auditory Stimulus when participants responded to auditory stimuli, $F_{(4,28)}=13.2$, $p<.005$. This interaction reflects the congruence effect also found in the RTs, because the least errors were made in congruent and the most errors were made in incongruent conditions.

3.1.3 Discussion

Clear congruence effects were found with abstract stimuli whose direction of movement was varied. Effects of semantic congruence already existed in the first block of the experiment. Large effects were observed especially for responses to auditory stimuli. As in the previous experiments, this might result from the larger difficulty of classifying auditory than visual stimuli. This hypothesis was further confirmed by more errors when responding to the auditory modality (7.9%) than when responding to the visual modality (3.8%).

In contrast to Maeda et al. (2004) who used ambiguous visual stimuli, I could show effects with a more objective measure. Despite clearly identifiable stimuli, congruence effects between vision and audition were discovered. Furthermore, I found effects on both target modalities, while Maeda et al. explored effects of auditory stimuli on the ambiguous visual stimulus only. This bidirectional result implies that modalities influence each other. A change in pitch seems to have its equivalence in direction of movement of a visual stimulus. Results do not depend on whether stimuli are ambiguous (Maeda et al.) or unambiguous.

The rising and falling pitch of the sound seems to be congruent to upwards and downwards movement of a visual stimulus. What would happen if the visual stimulus moves to the left or right? Is the association of pitch and disk-movement learned throughout the experiment? The results of the comparisons in the first block argue against a learning effect. Experiment 6 will test whether

there is a fixed correspondence between rising or falling pitch and upwards or downwards movement.

Introduction of a stationary visual stimulus was also new compared to Maeda et al. (2004). This stimulus was employed as a neutral condition, but congruence effects were also found between no change in pitch and no movement of the disk. Combining the stationary stimulus with a moving stimulus resulted in mean RT between those of congruent and incongruent conditions. Thus, stationary stimuli can be regarded as a neutral condition.

Maeda et al. (2004) have discussed the relevance of semantics for the effects. They reported a control experiment with spoken words (i.e. Japanese for ‘up’ and ‘down’), in which no effects on the perception of the visual stimulus were found. In their opinion this disproves the influence of semantics. This is further supported by another control experiment. Maeda et al. varied SOAs (-600 ms to +600 ms) and found the largest effect when vision precedes audition by approximately 50 ms. In their opinion, this argues against top-down influences, such as semantic priming, but in contrast it speaks for effects on a perceptual level, which the authors call ‘metaphorical congruence’. To my opinion, it remains a question of definition if it is called ‘semantic’ or ‘metaphorical’. As mentioned in 1.4, semantics is commonly understood as the meaning of a piece of information (Encyclopedia Britannica, 2002). In the experiments of Maeda et al. (2004) the context played some role, because the sounds had different effects and only differed in the direction of pitch-movement. It seems likely that a linguistic stimulus is not as congruent to an ambiguous abstract movement as an abstract tone. Thus, semantic influences might have been diminished in their experiments, because stimulus properties were incongruent. It seems doubtful that perception of the ambiguous stimulus can be affected by linguistic stimuli. If this hypothesis it true, congruence of stimulus properties seems also to play a role. Stimulus properties for auditory stimuli in the present experiment were amplitude, timbre, frequency range and direction of the frequency modulation. Color, form, size, speed and direction of the movement were stimulus properties of visual stimuli. The only properties that were not kept constant were direction of frequency modulation and direction of move-

ment of the disk. The meaning or semantics is the same for visual and auditory stimuli, that is, the (subjective) perception of the movement direction. Thus, stimulus properties differed in congruent conditions, whereas the semantics is the same. However, stimulus properties for visual and auditory stimuli seem to be congruent. Consequently, we cannot separate modulation of stimulus properties and modulation of semantics in the present experiment. Stimulus properties of linguistic auditory stimuli differ largely from the visual stimulus properties. Under these conditions, semantics on an abstract level is the same, but not on a perceptual level.

3.2 Experiment 6: Movement directions of rising and falling pitch

In Experiment 5, I found congruence effects of movement directions of visual and auditory stimuli. The movement direction of visual stimuli was varied on a vertical axis. It is unclear though, if auditory stimuli were perceived to move in the same directions. Do congruence effects also exist when the visual stimulus moves horizontally? The importance of this question is the following: Congruence between movement directions of auditory and visual stimuli might be developed throughout the experiment. This learning process is unlikely, because congruence effects were found as early as in block 1. But congruence might also be forced by the employed stimulus combinations. It is possible that congruence effects are smaller for a rising tone and an upwards moving disk, if different visual stimuli are presented, such as movement to the right.
The present experiment has the objective to find out, which direction of the visual stimulus is compatible to sounds with ascending and descending pitch. Therefore, eight movement directions of the disk will be included, as well as a stationary disk. Each of the nine visual stimuli is presented simultaneously with one of the three auditory stimuli from the previous experiment. The task is to indicate if the pitch of the auditory stimulus is rising, falling or remains constant. Responses to visual stimuli are not surveyed, because of stronger effects in this direction in Experiment 5.

3.2.1 Methods

Participants. Eight undergraduate students (one male) received course credit for participation. The average age was 26.1 years (range from 20 to 36). All were naive as to the purposes and hypotheses that motivated the study. They reported normal or corrected-to-normal vision, as well as normal hearing. All participants were right-handed.

Design. Auditory Stimulus consisted of rising, falling, or constant pitsch. Visual Stimulus had nine levels (eight movement directions plus no movement). Responses had to be made to auditory targets only.

Apparatus. The same soft- and hardware as well as the same room were used as in Experiment 1.

Stimuli. Visual stimuli consisted of a black disk with the same parameters as in Experiment 5. In each trial it moved in one of eight possible directions (see Figure 3.4), or it remained stationary. All other conditions were as in Experiment 5.

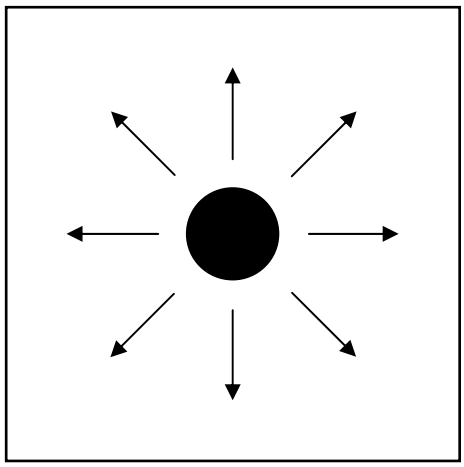


Figure 3.4: Movement directions of the visual stimulus in Experiment 6. Arrows indicate the eight implemented directions.

Task. Participants were instructed to indicate as fast and accurately as possible whether the sound frequency increased, decreased, or remained fixed. Throughout the whole experiment the visual stimulus was irrelevant for the task.

Procedure. Participants were tested in one 1-hour session. Six blocks with 162 trials each were presented. Thus, each stimulus combination occurred 36 times. The trial events were identical to Experiment 5, with the exceptions that participants always responded to auditory stimuli, and more movement directions of the visual stimulus were implemented.

Data analysis. RT and error data were treated as in Experiment 1.

3.2.2 Results

Response times. Mean RTs for each condition are illustrated in Figure 3.5. As in Experiment 5, the auditory stimulus influenced auditory judgments, which was revealed by a significant main effect of Auditory Stimulus, $F_{(2,16)}=8.1$, $p< .005$. Responses to downwards moving stimuli were in average as fast as to stationary auditory stimuli, except for the congruent condition for stationary stimuli, which led to much faster mean responses (539 ms for downwards movement and 523 ms for stationary sounds). Mean responses to upwards moving sounds were decelerated by the incongruent visual movements. Thus, mean responses were much slower than to the other two auditory conditions with a mean of 566 ms. Visual Stimulus also had a main effect, $F_{(8,64)}=4.2$, $p<.05$. This showed that visual direction influenced judgments of auditory pitch movement. This effect may be best explained by the significant interaction, $F_{(16,128)}=9.0$, $p<.001$. Responses to the upwards moving sound were facilitated when an upwards moving visual stimulus was presented simultaneously. This could be observed for straight upwards movement as well as for movement to the upper right and upper left. The corresponding case was found for downwards moving sounds. The course of the graph of stationary sounds was mostly horizontal except for a decline in RTs when a stationary (and thus congruent) visual stimulus was presented simultaneously. Summarized, RTs showed significant effects of semantic congruence.

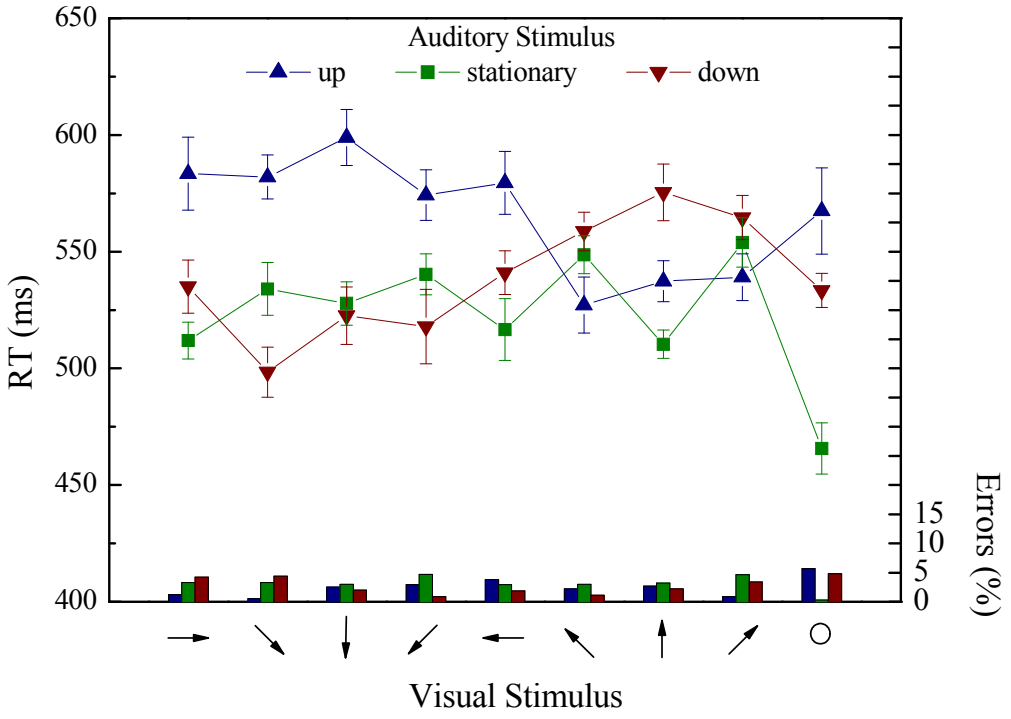


Figure 3.5: Mean RTs and errors of Auditory Stimulus as a function of Visual Stimulus of Experiment 6. The arrows on the asymptote indicate the direction of movement. No movement is symbolized by a circle.

Errors. On average, 2.8% of errors were made. Although the percentage was low, statistical analyses were computed. Main effects of Auditory Stimulus and Visual Stimulus were non-significant, $F_{(2,16)}=1.3$, $p=.29$ and $F_{(8,64)}=0.4$, $p=.84$ respectively. However, the interaction reached significance, $F_{(16,128)}=4.2$, $p<.005$.

3.2.3 Discussion

Results from Experiment 5 were supported in the present experiment. The pitch of the auditory stimulus seems to move upwards, downwards or remain stationary, as was supposed. For example, responses to the falling tone were fastest when it was accompanied by a disk that moved downwards. When the same sound was presented together with a disk that moved in an upwards direction, responses were decelerated. Horizontal movement seemed to be rather neutral with mean RTs between those of congruent and incongruent movement. The more different a visual movement compared to the perceived direction of the tone, the larger was the effect of incongruence.

Not only straight up and down movements enhanced RTs to the congruent tones. Movements to the diagonal upper-right and upper-left enhanced perception of the rising tone. Accordingly, movements to the lower-left and lower-right enhanced perception of falling tones. This indicates that falling and rising pitch are perceived as if moving in any upwards and any downwards direction respectively. The perceived direction of tones is not restricted to straight up or down movement. A horizontal bias, i.e. just to the upper-right but not upper-left, could have been possible due to a learned reading direction or other cognitive biases. However, movement to the upper-left and upper-right have similar effects.

3.3 Experiment 7: Detection task with simple stimuli

After showing congruence effects between perceived movements of visual and auditory stimuli, the question arises on which perceptual level integrating occurs. Experiments 3 and 4 already tried to discover whether semantic influences persist on a lower integration level. Results indicated no significant effects of semantic congruence. However, trendwise effects were evident. It was discussed that the stimuli were heterogeneous and variance of the responses may thus have been large. The stimuli of the present series of experiments are more homogeneous, which might thus help to solve the puzzle. If no effects of semantic congruence are found on a lower level task, this is a strong argument that integration of semantically congruent multimodal stimuli occurs at a rather late processing stage.

Maeda et al. (2004) also addressed the question of integration level. They reasoned that their effects originated on a perceptual level, and that top-down effects are unlikely. This argument is based on two of their control experiments.

First, Maeda and colleagues presented the spoken words ‘up’ and ‘down’ instead of rising and falling tones, which did not influence judgments of the ambiguous visual grating. The authors reasoned that the effects were not based solely on semantic processing, but instead, integration occurred on a perceptual level. As mentioned in 3.1.3, stimulus properties of linguistic stimuli were rather incompatible to the drifting grating. Perception of the ambiguous grating was not influenced by the linguistic stimuli. Participants regarded the spoken words as incompatible with the grating and did not respond in accordance with the meaning of the words. Does an effect on the perceptual level imply exclusion of semantic processing? Processing tones and spoken words is rather different (cf. 1.4). It is thus not surprising that no effect was found.

Another critical point is that only four Japanese and four non-Japanese participated. The latter ones did not understand the meaning of the auditory stimuli. Therefore, semantics could not have any effect, as the statistical power is too small to detect a small difference with a small sample size. Therefore, effects of semantics cannot be ruled out with these findings.

In the second control experiment, Maeda et al. implemented SOAs. Tones were separated by a variable time interval from presentation of the grating. Effects

were largest when the visual stimulus preceded the auditory stimulus by 50 ms. The authors argued that semantic priming was therefore unlikely. However, my experiments showed that semantic congruence affected responses even for simultaneously presented stimuli.

The present experiment tries to find out at which processing stage the effect occurred. To do so, a task is needed that does not require deep processing. In the previous experiment, participants had an advantage of attending to the irrelevant modality, because stimulus combinations were relevant for the task. To explore lower level processing a task is needed, for which congruence is irrelevant. Simple go/no-go detections of sounds as in Experiment 3 are a candidate. When participants respond as soon as they perceive a sound, no deeper processing is needed, and congruence is irrelevant for the task. If semantic congruence acts on a low perceptual level, congruence effects should also be found in this experiment. If, however, deeper processing is needed, RTs should not depend on congruence.

3.3.1 Methods

Participants. Eight undergraduate students (one male) received course credit for participation. The average age was 21.4 years (range from 20 to 29). All were naive as to the purposes and hypotheses that motivated the study. They reported normal or corrected-to-normal vision, as well as normal hearing. Every participant reported right-handedness.

Design. In contrast to Experiment 5, Target Modality was auditory throughout the experiment. Thus, participants only responded to auditory stimuli. Rising, falling, and constant tones were the same as in Experiment 5. In one fourth of trials, no auditory stimulus was presented. Visual Stimulus remained as in Experiment 5.

Apparatus and Stimuli. Hard- and software as well as visual and auditory stimuli were the same as in Experiment 5.

Task. Participants were instructed to respond as fast as possible when the auditory stimulus was presented. In this go/no-go simple detection task a key was to be pressed when a stimulus was detected (go). When no sound was presented, participants were to wait for the next trial (no-go).

Procedure. A total of 1800 trials were presented in two sessions with five blocks each. Blocks were separated by 30 sec breaks. Additional 24 practice trials per session were excluded from analysis. Trial events were as in Experiment 5, except that participants had a maximum of 1500 ms to give a response. After a random intertrialinterval between 1000 ms and 1500 ms, the next trial started automatically.

Data analysis. RT data were treated as in Experiment 1. Error data were not analyzed, because only about 1% incorrect responses were made.

3.3.2 Results

Mean RTs for the combinations of visual and auditory stimuli are reported in Figure 3.6. All mean RTs were between 280 ms and 288 ms. Thus, responses were very fast and differences small. Accordingly, main effects and interaction of the ANOVA were non-significant. Levels of Visual Stimulus as well as Levels of Auditory Stimulus did not affect perception differently, $F_{(2,14)}=0.7$, $p=.47$ and $F_{(2,14)}=0.7$, $p=.43$ respectively. An interaction of the factors was not evident, $F_{(4,28)}=0.8$, $p=.54$, indicating that congruence did not influence responses in a detection task. Furthermore, trendwise differences were not in accordance to the hypothesis of semantic processing.

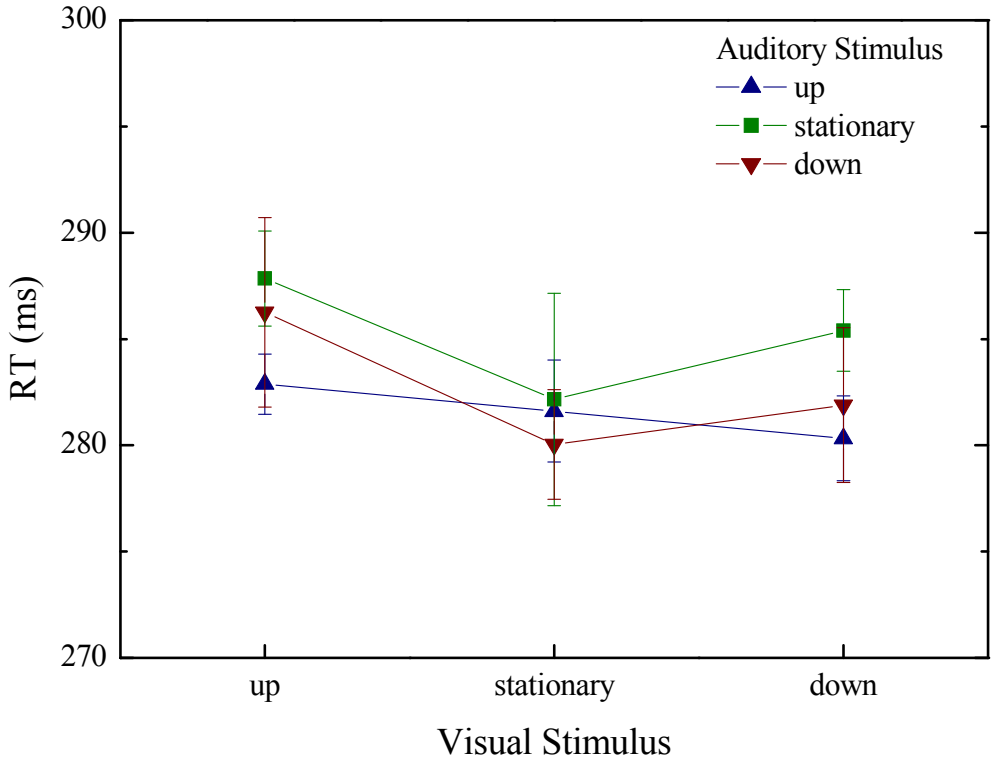


Figure 3.6: Mean RTs of Auditory Stimulus as a function of Visual Stimulus in Experiment 7.

3.3.3 Discussion

No effects of semantic congruence were evident in this task requiring low level processing. This confirmed the hypothesis that semantic processing of audiovisual stimuli occurs at a later processing stage. Deeper processing was not necessary to perform the task. How are the results to be interpreted? Maeda et al. (2004) observed effects of the same auditory stimuli as here on the perception of an ambiguous grating. They argued that these effects occurred on a perceptual level and that semantics was irrelevant. Does this conclusion stand in contrast to the present findings of the absence of low level effects? Maeda et al. used a choice reaction task (i.e. report of perceived direction), which was easier to perform with auditory stimuli. I used instead a go/no-go detection task, which made semantics of the stimuli irrelevant. Thus, processing levels of the stimuli probably differed. Therefore, Maeda's results can be explained by assuming that semantics can affect perception only when information is

processed semantically. The difference in the target modality might further explain why Maeda et al. found effects, while effects were absent in my experiment. Maeda's group found effects from audition onto vision. Conversely, the present experiment focused on effects from vision onto audition. Previously, I found larger effects in this direction. The importance of semantic processing is therefore supported because no congruence effects were found even in this previously dominant direction.

3.4 Experiment 8: Effects of environmental sounds on the path of disks

When two identical objects move towards each other, coincide and then move apart again, they may be perceived as streaming through or bouncing off each other (Metzger, 1934). The perception of this ambiguous display is strikingly influenced by a sound. Sekuler et al. (1997) presented a brief click at the coincidence and participants perceived bouncing objects (i.e. disks) in about 62% of trials (about 80% in Watanabe and Shimojo, 2001a). In contrast, only about 22% perceived bounces in trials without the sound. See Figure 3.8 (lower panel) for an illustration of the two perceptions. This effect is supported by functional imaging. Bushara et al. (2002) found increased activity in multimodal areas (e.g. SC and insula) and decreased activity in unimodal areas (e.g. superior temporal gyrus and medial occipital cortex).

This intersensory influence of an auditory stimulus on visual perception (cf. 1.1) may have been learned in the real world as most collisions are accompanied by a simultaneous sound and an intersensory association is thus learned (Shams, Kamitani & Shimojo, 2004). Further experimental findings revealed that a synchronized sound embedded between two flanker sounds, which are identical to the synchronized sound, attenuates the bounce inducing effect. When the embedded sound differs from the flanker sounds, bounce perceptions dominate again. The interpretation of the authors was that the sound needs to be salient to induce bouncing (Watanabe & Shimojo, 2001a). Is salience the only required condition to increase perceived bounces? Besides the effect of a sound, intramodal manipulations may have similar effects, as found by Sekuler and Sekuler (1999). A short pause of the movement, disappearance of the objects and deceleration of the movement led to increased bounce perceptions. Furthermore, a nearby moving irrelevant object could influence perception according to its direction of movement. When the irrelevant object changed the direction synchronously with the coincidence, participants saw bounces, whereas they saw streams, when the irrelevant object moved consistent with a streaming disk (Kawachi & Gyoba, 2006). To sum it up, salience may not be the only relevant condition to induce bouncing.

Other stimuli may also lead to more bounce perceptions. A flash as well as a brief vibration have a similar effect (Watanabe & Shimojo, 1998; 2001b). An effective time window was found similarly for each of the three modalities and

ranged from -300 to +200 ms for an auditory stimulus, from -100 to +100 ms for a visual and -600 to +100 ms for a tactile stimulus (Shimojo & Shams, 2001; Remijn, Ito & Nakajima, 2004). This indicated that the stimulus needed to be temporally synchronized with the coincidence of the objects and that attention might play an important role. All presented stimuli and manipulations of movement can be regarded as attracting attention. This is further stated by the fact that an occluded coincidence did not induce bounces, in contrast to a disappearance of the objects for the same time. The occluder was seen before and did thus not attract attention as much as a sudden disappearance without occluder.

What else could have an impact on perception of the ambiguous motion display? If the ecological relevance stated by Shams et al. (2004) is correct, the association between sound and display is of importance. Shams et al. (2004) further referred to personal communication between Watanabe and Shimojo, who described that the "sound has to have a sharp onset to induce this effect" (p. 28). But why does a slowly rising sound not induce bouncing? Just because of the physical characteristics? It seems plausible that semantics play a role. Sounds that represent bounces are, for example, colliding cars or bouncing billiard balls. Both sounds have rather sharp onsets. A stream sound however has rising amplitude as two trains passing each other in opposite directions or an airplane flying by. In other words, the semantic content may be important.

The present experiment tries to explore, whether sounds that represent bounce induce visual perception of bouncing disks, whereas sounds that rather represent streams induce seeing streams. Therefore, two disks move towards each other, coincide behind an occluder and moved apart again. By using two different colored disks (i.e. blue and red) the path of motion can be determined. Thereby, congruence between sounds and the path of the disks can be varied. Furthermore RTs can be used to measure the bounce inducing effect, as has been shown previously by Sanabria, Correa, Lupiáñez and Spence (2004). To compare the results to other stream-bounce experiments, ambiguous displays (i.e. both disks had the same color) are used in a third of trials. The bounce sound was recorded from colliding billiard balls. To get a stream sound, the bounce sound has been reversed and thus sounds like a steam (cf. methods). In addition, in a third of trials no sound is presented. It is hypothesized that the

stream sound is similar to the no sound condition, because both are semantically congruent to steaming disks. In contrast, the sharp bounce sound should be congruent to bouncing disks and induce perception of bounces. Effects for the sounds should be found for RTs in unambiguous paths and reported perceptions in ambiguous trials.

3.4.1 Methods

Participants. Twelve undergraduate students (three male) received course credit for participation. The average age was 23.9 years (range: 19 to 34 years). All were naive as to the purposes and hypotheses that motivated the study. They reported normal or corrected-to-normal vision, as well as normal hearing. Six participants reported right-handedness.

Apparatus. The apparatus remained the same as in Experiment 1.

Stimuli. A fixation cross (width/height: 0.3°) was visible 0.6° below the center of a background square throughout all trials. The background square in the center of the screen was light grey and a width and height of 9.8°. Red and blue disks with a size of 15 pixels (0.4°) moved with steps of 3 pixels at a speed of 6.5° per second in opposite directions (Figure 3.8 and 3.9). The disks started from the outer bounds of the background square, 2.1° above fixation cross. The coincidence was covered by a dark grey occluder (height: 5.5°; width: 1.1°). The disks disappeared behind the occluder for 238 ms. A black disk (diameter: 0.4°) underlaid the fixation cross to indicate that a response was entered. Errors were signaled by a white X (size: 1.4°) in the center of the background square.
Auditory stimuli were two sounds of 173 ms duration. The bounce sound, taken from a free internet source (www.findsounds.com) was recorded from two colliding billiard pool balls. The stream-sound was the same sound, just reversed in time. Figure 3.7 shows the waveforms of the two sounds. The hard versus the soft onset define the bounce sound versus the stream sound.

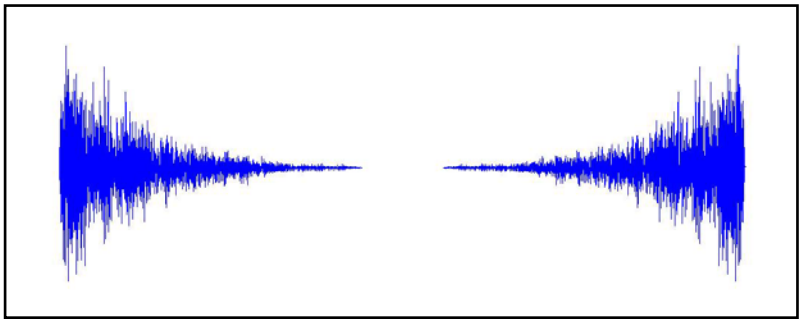


Figure 3.7: Waveforms of sounds representing bounces (left) and the reversed sound for streams (right).

Design. The path of the disks was either unambiguous or ambiguous. Figure 3.8 illustrates all implemented paths. Unambiguous and ambiguous trials were analyzed separately (RTs vs. reported perception). But they were presented in random order. Two third of trials were unambiguous and one third was ambiguous. Auditory Stimulus with the levels stream, bounce and no sound was the same for both paths.

Half of the unambiguous paths (different colored disks) were bounces and half streams. Thus, Visual Path consisted of the levels stream and bounce. RTs and errors were the dependent variables in the unambiguous trials. In the ambiguous trials, reported perceptions served as the dependent variable.

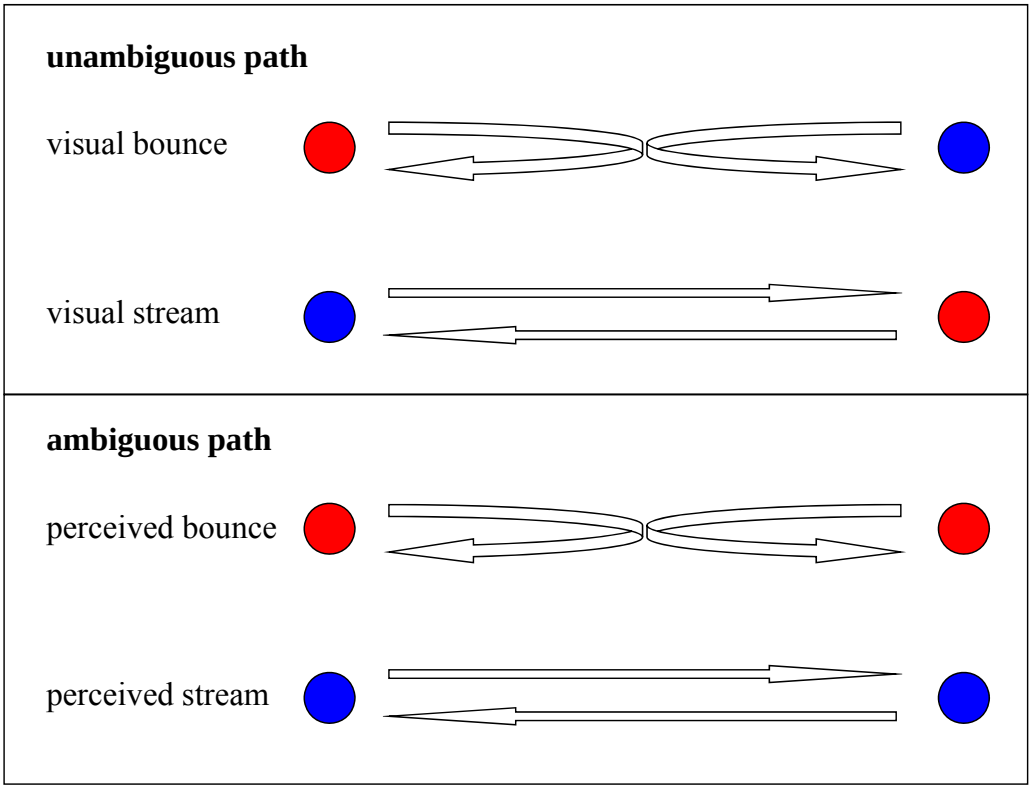


Figure 3.8: Visual stimuli of Experiment 8. The path was unambiguous by presenting a blue and a red disk (balanced starting positions) that moved towards another and either moved back to their starting position (visual bounce) or continued to move to the other side (visual stream). Two disks of the same color were presented in the ambiguous path condition. The movement direction (bounce vs. stream) depended on reports of participants' perceptions.

Task. Participants were instructed to respond as fast and accurately as possible and indicate whether they saw the two disks as steaming through or bouncing off each other. Participants were to observe sound and vision but to respond to vision only. Sounds were irrelevant for the task.

Procedure. Participants were tested in a one-hour session. Instructions were given onscreen with practice trials that were excluded from the analyses. The experiment contained three blocks with one-minute breaks after every block. The trial order was randomized within each block. 216 trials were presented in a block, thus, every participant completed 648 trials.

An example of trial events is shown in Figure 3.9. Each trial started with presenting a fixation cross for 1000 ms. Afterwards, the two disks appeared on the sides of a background square and started moving towards one another. In the center, the disks disappeared behind an occluder, coincided and appeared again

moving in the same direction as before (stream) or in the direction where they came from (bounce). The sound was presented simultaneously with the coinciding disks. When participants saw the disks after reappearance, they could give their response. After making a response, the fixation cross was underlaid by a black disk, to indicate that a response was entered. In the unambiguous trials, errors were signaled by presenting a white X for 500 ms. The next trial started automatically.

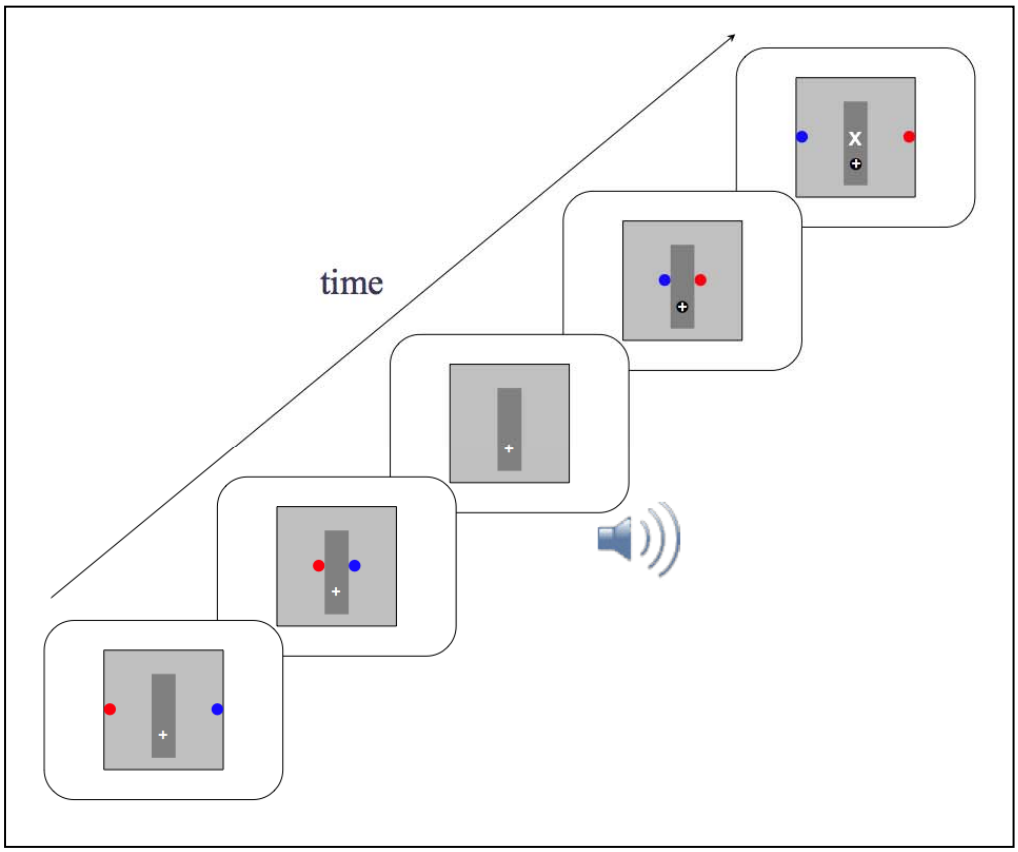


Figure 3.9: Example trial for a visual stream from Experiment 8. A red and a blue disk move towards each other, coincide behind an occluder and move apart again. A sound is presented simultaneously with the coincidence.

Data analysis. RT and error data were treated as in Experiment 1.

3.4.2 Results

Unambiguous visual path.

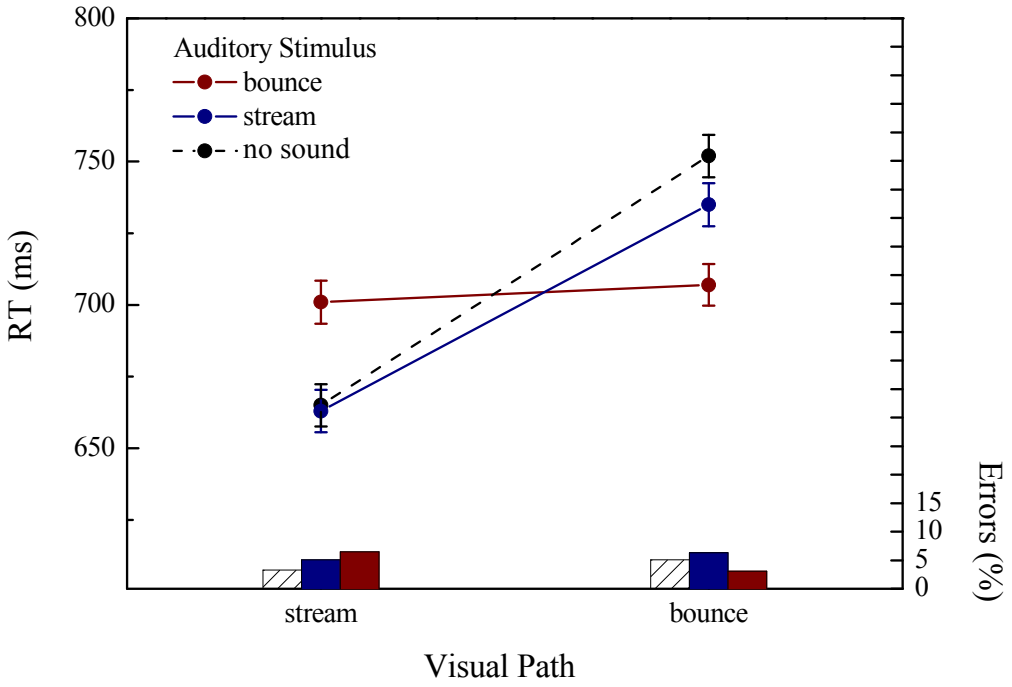


Figure 3.10: Mean RTs (lines) and errors (columns) of Auditory Stimulus as a function of Visual Path.

Response times. Results of unambiguous trials were illustrated in Figure 3.10. Responses to stream trials were faster than responses to bounce trials. This was supported by a significant main effect of Visual Path, $F_{(1,11)}=33.4$, $p< .001$. The auditory stimulus alone had no significant effect [$F_{(2,22)}=0.67$, $p= .52$], but the interaction with Visual Path was significant, $F_{(2,22)}=19.2$, $p< .001$. As can be seen in Figure 3.10, responses were fastest when sound and sight were congruent. For example, auditory and visual bounces led to faster responses than auditory bounce and visual stream. The stream sound and no sound condition led to very similar responses and both had strong congruence effects. The bounce sound did not differ much in both Visual Path conditions.

Errors. In the unambiguous trials, participants made 4.9 % errors on average. Thus differences were rather small. Nevertheless, an ANOVA with the arcsine transformed errors was calculated. Neither Visual Stimulus nor Auditory Stimulus had a significant effect on the error rates, $F_{(1,11)}=0.2$, $p= .67$ and

$F_{(2,22)}$=1.2, p= .32 respectively. The significant interaction [$F_{(2,22)}$=5.7, p< .05] indicates that errors were less, when Auditory Stimulus and Visual Path were congruent than when they were incongruent (see columns in Figure 3.10).

Ambiguous visual path.

When both disks had the same color, the frequency of reported sensations was the dependent variable. Therefore, data were arcsine transformed and an ANOVA for Auditory Stimulus was calculated. Results indicate that the sound significantly influenced the reported perception, as indicated by a significant main effect of Auditory Stimulus, $F_{(2,22)}$=31.9, p< .001. Figure 3.11 illustrates the influence of sounds on the perceived direction of the disks. As in the results for the unambiguous visual path, no sound and stream sound condition produced similar results.

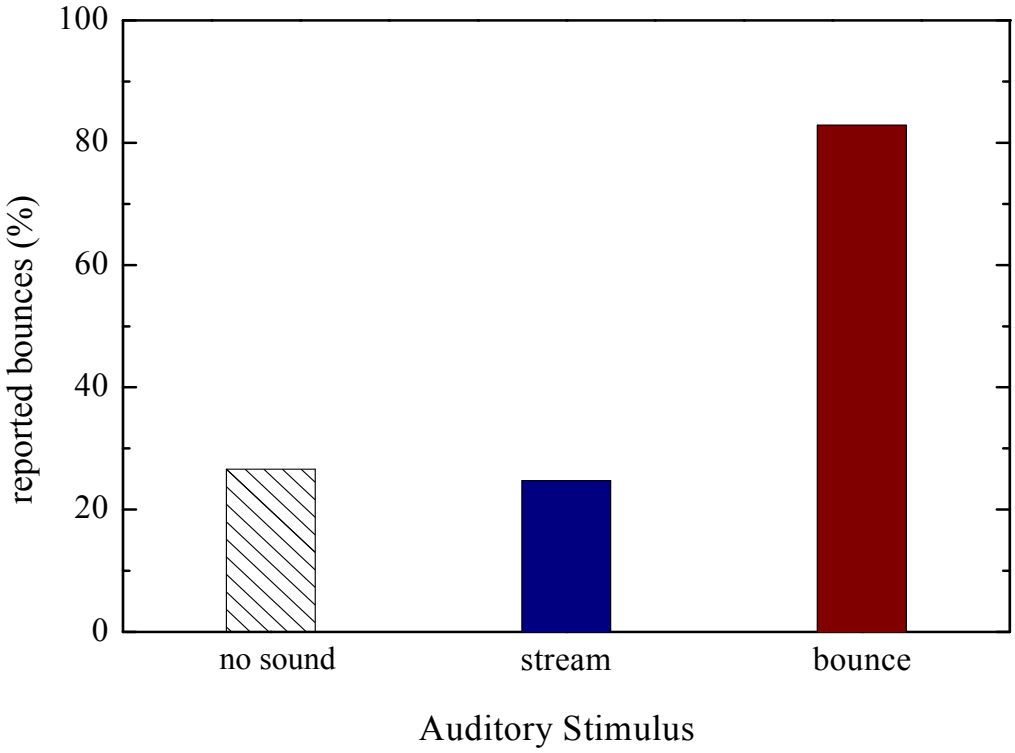


Figure 3.11: Percentage of reported bounces of Auditory Stimuli in ambiguous trails of Experiment 8.

3.4.3 Discussion

Effects of semantical sounds were found in the speeded responses as well as in reports of perceived paths. Responses were faster to bouncing disks when accompanied by a bounce sound, as compared to stream sound or no sound. Accordingly, in ambiguous trials, the bounce sound induced perception of visual bounces, and vice versa for stream sounds and no sounds. Responses to stream sounds and no sounds did not differ in speed. Thus, it could be demonstrated that salient sounds can also evoke stream perception, similar as the no sound condition, or the identical embedded sounds in previous experiments (Watanabe & Shimojo, 2001a; Sekuler & Sekuler, 1999). Salience of the sound did not play a role here, because both sounds were not embedded in other sounds.

Sounds were the same except for their temporal order. However, they had opposite effects. Semantic congruence is a possible explanation, but is it the only one? What is the role of the physical characteristics in the current findings? The sounds differed in the rise and fall of amplitude (see Figure 3.7). The sudden onset of the bounce sound and the rising stream sound may be processed differently. Differences may arise from several possibilities. First, the perceptual-centers (p-centers) of the sounds are different for the sounds. A p-center indicates the perceived temporal occurrence of an auditory stimulus (Morton et al., 1976). Thus, in order to perceive two auditory events as simultaneous, not their onsets must be simultaneous but their p-centers. P-centers largely depend on the rise times (Scott, 1998). Hence, the p-center for the bounce sound is earlier than the p-center of the stream sound. Different p-centers might result in different asynchronies between the respective sound and the visual event. Nonetheless, the sounds were 173 ms long and the disks were 238 ms behind the occluder. According to the effective time window described in Experiment 2, both p-centers lay within this window and should thus evoke bounce perceptions equally often.

Second, the loudness of rising tones is often overestimated, but underestimated for falling sounds (Neuhoff, 1998). The ecological validity of this result is that rising tones reflect movement towards somebody, whereas falling tones reflect movement away. Assigned on the current findings, the stream sound should

attract more attention than the bounce sound. Thus, the stream sound should induce the perception of bouncing disks and the bounce sound should accordingly induce perception of streaming disks. The data point in the opposite direction, which strongly argues against this hypothesis.

As the semantics of the sounds correlates with the physical properties of the sounds, it is not possible to dissociate between them.

These accounts are also discussed in a developmental study of Scheier, Lewkowicz and Shimojo (2003). They tested 4-, 6- and 8-months old infants with the ambiguous motion display, and presented tones during a habituation phase asynchronously and in a test trial synchronously, and vice versa. Older infants (6- and 8-months) responded differently to the habituated trials than to the test trials, indicating that they perceived a difference. This was not the case for 4-months old infants. The authors discussed two assumptions for this result. Either lacking experience with such stimuli or attentional mechanisms were not sufficiently developed in younger infants. Both assumptions reflect the current discussion.

According to Wallace (2004), postnatal sensory experiences play an important role in the development of multisensory integration. He discovered that neurons in the SC of cats, who were raised without any visual cues, lacked the typical response enhancement for multisensory stimuli. This result stresses the importance of experiences for the ability to integrate sensory information.

4. Semantic congruence of linguistic stimuli

4.1 Experiment 9: Semantic relation vs. response-congruence

One main objective of the current work was to explore semantic influences in nonlinguistic stimuli. However, nonlinguistic stimuli have some methodical constraints. For example, sounds of semantically related items can often not be distinguished. ‘Hammer’ and ‘nail’ would have similar associated sounds. Pictures of juice and nectar are also not distinguishable. In contrast, linguistic stimuli have the advantage that a larger pool of appropriate objects is available. Especially there are often no characteristic sounds for objects. A ear or a cake, for example, can hardly be presented as sounds.

Therefore, Experiments 9 and 10 explore semantic effects by means of linguistic stimuli, in order to expand the findings for nonlinguistic stimuli. As mentioned in 1.4.1, there is frequent research on crossmodal speech processing. Two paradigms may be distinguished. First, studies exploring differences between lip/face reading and spoken words, a domain usually cited to as speech reading. Second, there are studies implementing written and spoken words, usually referred to as crossmodal priming.

The experiment of Calvert at al. (2000) is an example that found effects of semantic relation on speech reading. One disadvantage of this study was that differences in semantics went along with temporal differences, i.e. when different words were spoken visually and auditory, this was accompanied by a different timing of lip movements. Dubbed movies are an illustration for this effect. This problem is avoided when using written and spoken words.

Holcomb and Neville (1990) have been one of the first to compare intramodal visual and auditory semantic priming. They found faster responses and fewer errors for semantically related, as compared to unrelated words within both modalities. Furthermore, an analysis of ERPs resulted in a difference in the amplitude of the N400 for related and unrelated words. The N400 effect was larger and in a wider range of SOAs (200 to 800 ms) for the auditory than for the visual modality (300 to 600 ms).

Holcomb and Anderson (1993) expanded these results to crossmodal semantic priming. They primed crossmodally in both directions, that is, from vision onto audition and vice versa, with SOAs between 0, 200 and 800 ms. The authors found semantic priming effects in both directions and at all SOAs in a speeded lexical decision task (i.e. word or non-word). Overall, effects were larger from vision onto audition than in the other direction. Within this direction, effects increased with increasing SOA, whereas for audition onto vision, priming effects were smallest at a SOA of 200 ms. Additionally, ERPs were measured. The main result was that the N400 was more negative for unrelated than for related words. This effect was larger for auditory than for visual targets. Taken altogether, larger semantic priming effects were found from vision onto audition. This is concordant with the results from the present experiments.

The question arises whether the results from Holcomb and Anderson (1993) are comparable to the results of the present experiments. One main difference is the task. Holcomb and Anderson used lexical decisions while I used a categorization task. I did not find semantic effects when a different task was implemented (e.g. detection or TOJ). To compare effects of linguistic and nonlinguistic stimuli, the same task must be used. Therefore, an experiment with words from a living and a nonliving category will be conducted. Building up on Holcomb and Anderson (1993), targets will be always presented auditorily and primes visually with four SOAs (0, 100, 700, 800 ms). To separate semantic effects from category effects, semantically congruent, response-congruent and incongruent stimuli are employed. It is predicted that semantic influences will be found at all SOAs.

4.1.1 Methods

Participants. Eight undergraduate students (four male) received course credit for participation. The average age was 22.4 years (range from 20 to 27). All were naive as to the purposes and hypotheses that motivated the study. They reported normal or corrected-to-normal vision, as well as normal hearing. All participants were right-handed.

Apparatus. The room as well as hard- and software remained the same as in Experiment 1.

Stimuli. All target words were presented auditory. Forty-two German target words were recorded in a female voice. The target words were divided in two categories, i.e. living and nonliving. Furthermore they were subdivided in three subgroups each. The category Living consisted of humans (e.g. brother), animals (e.g. tiger), and body parts (e.g. finger). Subgroups of the category Nonliving were house holding items (e.g. table), food (e.g. bread), and clothing (e.g. shirt). They were clearly spoken in approximately the same speed with a mean duration of 666 ms. Targets were on average 5 letters long. No significant difference in duration of the spoken words between categories was found.
All primes were presented visually, i.e. in printed letters, in Arial, size 50. This resulted in a height of approximately 1.4°. The length of prime words was 5 letters on average. The number of letters for primes did not differ significantly between categories. The whole set of the original German primes and targets is included in the appendix.
The fixation cross had a size of 0.7°. Errors were signaled by a red X (height: 0.7°).

Design. The two factors in this experiment were Congruence Type and SOA. Congruence Type had three levels, i.e. response-congruent and semantically related, response-congruent and semantically unrelated, as well as response-incongruent. Examples for response-congruent and semantically related words are the prime ‘sister’ for the target ‘brother’. Primes were taken from the same subgroup as the target. Response-congruent and semantically unrelated words originated from different subgroups (e.g. the prime ‘bumblebee’ for the target ‘brother’). The third type was response-incongruent (e.g. the prime ‘television’ for the target ‘brother’).
SOA was varied in four levels, 0 ms, 100 ms, 700 ms, and 800 ms. All conditions were repeated-measures factors. RT was the dependent variable.

Task. Participants were instructed to classify auditory targets as living or nonliving. They were to press the appropriate key as fast and as accurately as poss-

ible. Written words were to be focused, but participants were instructed that written words were irrelevant to perform the task correctly.

Procedure. Four blocks with 126 trials each were presented in a one-hour session. Each target was presented three times in a block, once with every prime type. The order of SOAs was randomized for every participant so that throughout a session every prime-target combination was presented once at every SOA. This resulted in a total of 504 trials. Instructions were read to participants and sample trials were excluded from calculations.

A typical trial is illustrated in Figure 4.1. Each trial started with a fixation cross for 500 ms. Then, the prime was presented visually for 50 ms, followed by the auditory target after the corresponding SOA. Errors were signaled by presenting a red X for 500 ms. The intertrial-interval was varied randomly between 1000 and 1500 ms. The next trial started automatically.

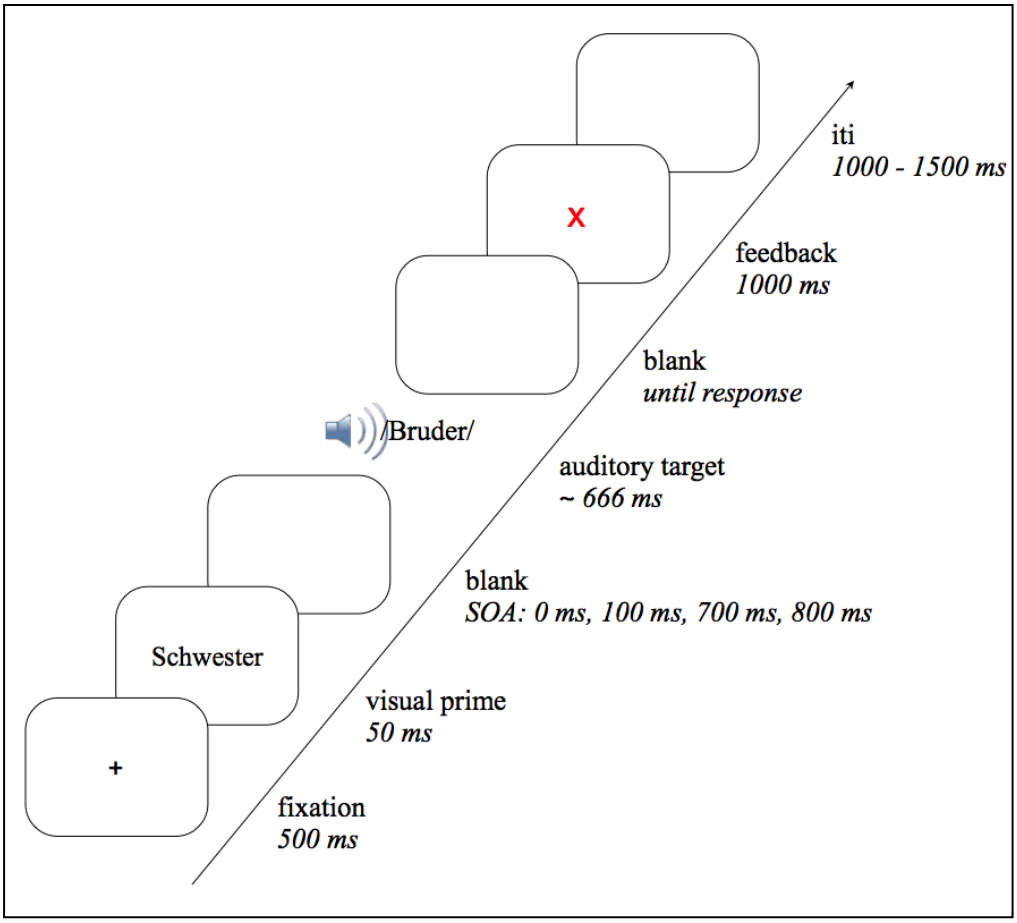


Figure 4.1: Trial events of Experiment 9 in an example. A prime word (here: sister) was presented visually. After a variable SOA, a target was presented auditory (here: brother). Incorrect responses were followed by feedback. The intertribal-interval varied randomly from trial to trial.

Data analysis. Overall conditions about 4.1 % errors were made. Error data were excluded from any analysis.

4.1.2 Results

Figure 4.2 illustrates the results of Experiment 9. Response-congruent and related primes led to the fastest responses at all SOAs. Response-congruent and unrelated primes led to slightly faster responses than incongruent primes. This is supported by a main effect of Congruence Type, $F_{(2,14)}=7.4$, $p< .05$.

RTs decreased slightly with increasing SOA. However, the main effect of SOA failed to reach significance, $F_{(3,21)}=3.1$, $p= .11$. Therefore, SOA did not influ-

ence responses. An interaction was also not evident, $F_{(6,42)}=0.7$, $p= .54$, indicating that Congruence Type was the only factor affecting RTs.

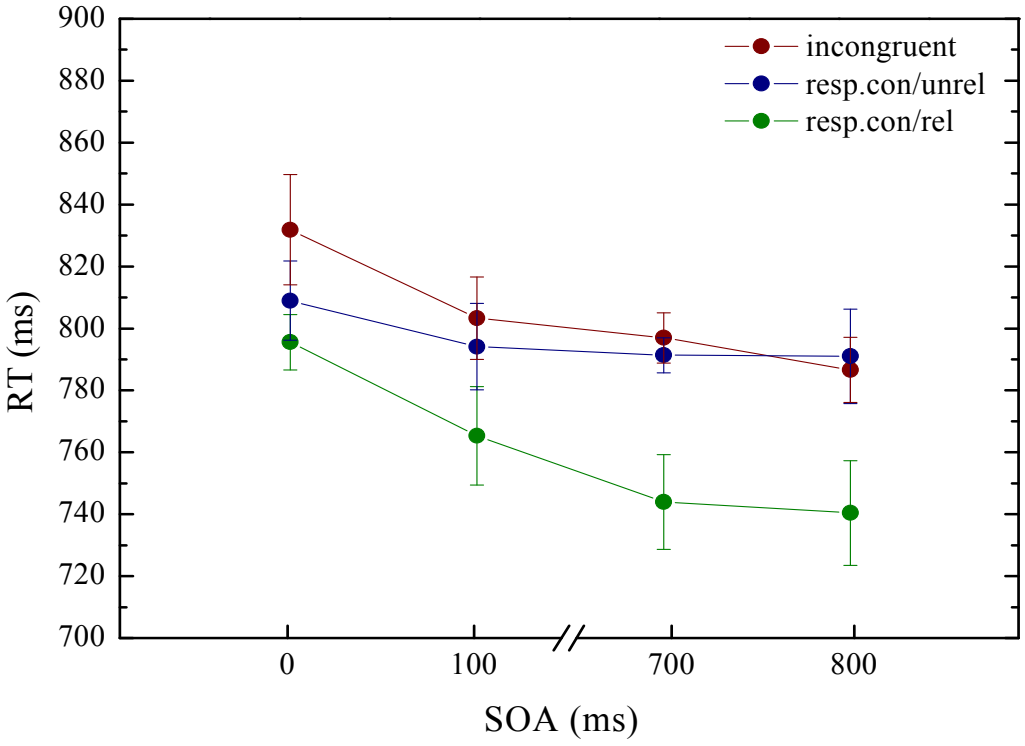


Figure 4.2: RTs of Congruence Type levels as a function of SOA in Experiment 9. Lines represent incongruent, response-congruent and unrelated (resp.con/unrel) as well as response-congruent and related primes (resp.con/rel).

To explore semantic priming and category priming, the net priming effects were calculated. Semantic priming is defined by the RT difference of unrelated and related response-congruent trials. RTs to incongruent trials subtracted from the mean RT of response-congruent primes (mean of unrelated and related) resulted in net category priming. Figure 4.3 shows that semantic priming increased with SOA. Category priming remained relatively constant over SOAs. Semantic priming was greater than category priming for SOAs of 100 ms and above. Semantic relation (e.g. mother - father) evoked greater enhancement than same response categories (e.g. calf - father).

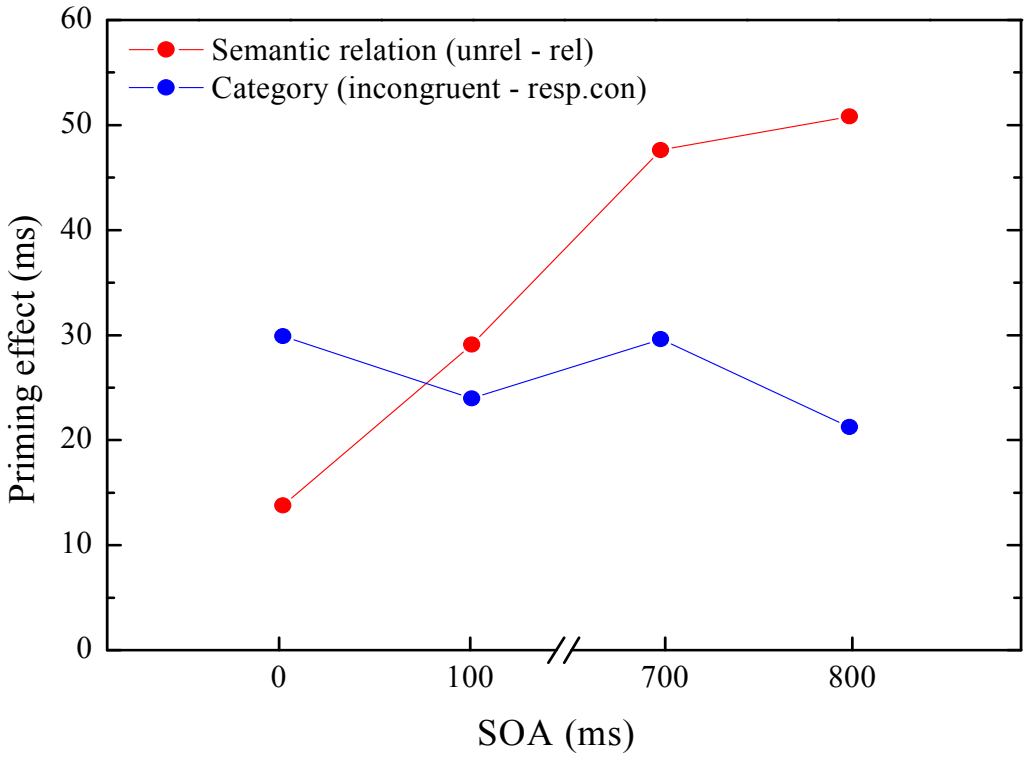


Figure 4.3: Net priming effects of semantic and category relations as a function of SOA. Semantic effects were computed by the difference in RTs between related and unrelated primes (unrel - rel), within the response-congruent condition. The difference between incongruent and response-congruent (resp.con) conditions estimates the category effects.

4.1.3 Discussion

Semantic effects were found over all SOAs, but effects increased strongly with SOA. This result was concordant with Holcomb and Anderson (1993). Interestingly, effects of response-congruence remained relatively constant over SOAs. That semantic congruence increases with SOA can be interpreted as a strategic effect (cf. Experiment 2). Importantly, already at an SOA of 100 ms, semantic congruence effects were as large as 30 ms. Thus, strategic effects cannot solely explain the results.

In contrast to the previous nonlinguistic experiments, the present experiment focused on semantic relation rather than stimulus-congruence. Does this imply that the previous experiments rather explored repetition priming? It is supposed

that the difference between sounds and pictures is so large that it is not comparable to repetition of written and spoken words. Orthographic and phonological properties of words are highly concordant (Holcomb, Anderson & Grainger, 2005), whereas pictures and sounds differ in several properties. The effects of stimulus-congruence or repetition priming in linguistic stimuli were explored in the following experiment.

How is semantic information processed by the brain? Holcomb and Anderson (1993) argued that crossmodal stimuli are first processed modality specific, but are integrated relatively early. Their and my results suggest that semantic information is rather amodal. A common semantic system seems to be responsible for processing information from different senses. A competing model would suggest conversion between the modalities. This would imply effects of semantic relation just after sufficient processing time. However, effects were found at all SOAs, not just at a SOA of 800 ms.

4.2 Experiment 10: Semantic relation in incongruent stimuli and stimulus-congruence

Effects of semantic congruence were explored in Experiment 9. Semantics influenced responses particularly at larger SOAs. The previous experiment included only positive SOAs, that is, a prime word preceding the target word. As in experiments 1 to 4, auditory stimuli were dynamic while visual stimuli were static. An open question is whether written words are perceived faster than spoken words. This would contradict the results of Experiment 4, which showed that pictures and environmental sounds are perceived as simultaneous when the visual stimulus precedes the auditory stimulus by about 40 ms. According to the temporal rule, perceived simultaneity may influence effect sizes (Stein & Meredith, 1993). To explore a wider range of SOAs, SOAs were now varied from -700 ms to +700 ms.

Experiment 9 had no baseline condition. Therefore, no prime word was presented in one sixth of all trials in the present experiment. Additional congruence conditions were furthermore implemented. Experiment 9 found facilitated responses to semantically related and response-congruent words compared to unrelated and response-congruent words. Do effects of semantic relation also exist for response-incongruent words? To explore this issue, semantic relation for response-incongruent words was varied. For example, the target ‘candy’ was presented along with the related word ‘child’ and the unrelated word ‘bird’, which were both response-incongruent.

Stimulus-congruent or repetition primes were also introduced. In the previous nonlinguistic experiments there were stimulus-congruent conditions. The compatible condition in the present linguistic experiments would be a repetition prime. However, orthography in written words and phonology in spoken words is quite concordant (Holcomb et al., 2005). Thus, stimulus-congruent words should facilitate responses much more than semantic congruent words. This is also tested in the present experiment.

The present experiment thus tries to expand and verify the results of Experiment 9. Negative SOAs, incongruent but related primes, a no-prime condition and a repetition primes-condition are implemented. The word pool is adapted and improved (some of the words of Experiment 9 were excluded). Semantic

effects are predicted to occur as before. Semantics should also have an effect on incongruent words. Responses to repeated words are probably faster than to all other congruence conditions, due to the high concordance of written and spoken words. Larger effects are expected with increasing SOA. Primes following the target (negative SOAs) are likely to interfere with the response. Thus, RTs increase and effects of semantic relation are smaller.

4.2.1 Methods

Participants. Eight undergraduate students (one male) received course credit for participation. The average age was 25.2 years (range from 19 to 40). All were naive as to the purposes and hypotheses that motivated the study. They reported normal or corrected-to-normal vision, as well as normal hearing. Six participants reported right-handedness.

Apparatus and Task. The apparatus and the participants' task were identical to that of Experiment 9.

Stimuli. Target words were presented auditorily. Thirty-six of the targets in Experiment 9 were the same. The rest was excluded (one word in every subcategory), because of no response-incongruent and semantically related primes could be generated. Some words of Experiment 9 were exchanged to improve the set of words. For example, the response-congruent and unrelated prime for the target 'hunter' was 'cattle' in Experiment 9, which seems not completely unrelated. In Experiment 10 this prime was changed to 'jellyfish', which is less related to 'hunter'. See the appendix for a complete list of all prime and target words.

Design. The factor Congruence Type had six levels. As in Experiment 9 targets were preceded or followed by response-congruent and semantically related primes (e.g. 'horse' and 'pony'), with response-congruent and semantically unrelated primes (e.g. 'horse' and 'nephew'), as well as response-incongruent and semantically unrelated primes (e.g. 'horse' and 'bottle'). In addition to

those of Experiment 9, response-incongruent and semantically related primes (e.g. ‘horse’ and ‘saddle’) as well as repetition primes (e.g. ‘horse’ and ‘horse’) were included. Targets were also presented without a prime. The range of SOAs was expanded with the seven levels -700, -500, -100, 0, +100, +500, +700 ms. Negative values correspond to primes (written words) presented before targets (spoken words), whereas positive values indicate presentation of targets before primes. In the no prime condition, accordingly SOA could not be defined. This level was still presented in the same frequency as the other levels of Congruence Type.

Procedure. 1512 trials were portioned in seven blocks and two approximately one-hour sessions. In each block every target was combined with every level of Congruence Type, resulting in 216 trials per block. In session 1, three blocks were preceded by a practice block, which was excluded from analyses. In session 2, four blocks followed 20 practice trials. In every block, a prime-target combination was presented with a different SOA. The assignment was constructed for every participant in a randomized form. Trial orders were randomized in every block. Trial events were as in Experiment 9.

Data analysis. Error data were not analyzed, because on average only 2.0% incorrect responses were given.

4.2.2 Results

Response times. Mean RTs are illustrated as a function of SOA and Congruence Type in Figure 4.4. The ‘no prime’ condition served as a descriptive baseline, and was excluded from calculations.

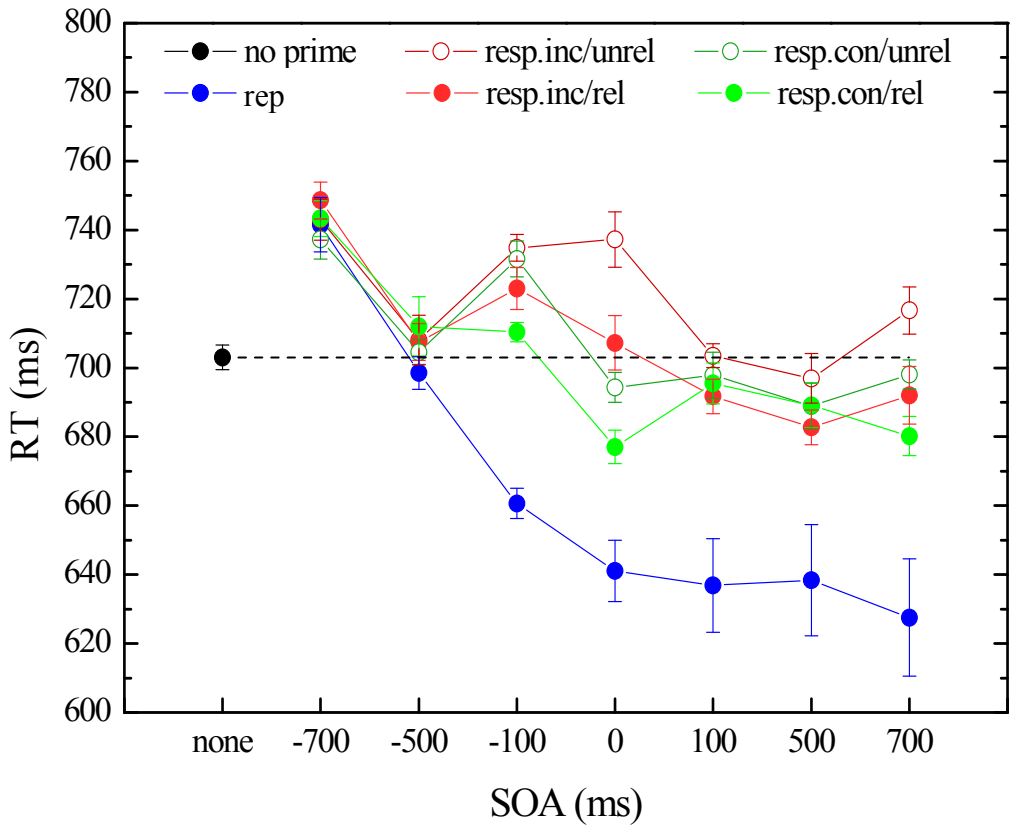


Figure 4.4: Mean RTs of Congruence Type as a function of SOA in Experiment 10. No prime-condition was used as the baseline (black). Response-incongruent primes (resp.inc) are shown in red, response-congruent primes (resp.con) in green and repetition primes (rep) in blue. Semantically unrelated primes (unrel) are visualized with darker colors and open dots, while semantically related primes (rel) have lighter colors and solid dots.

A repeated measures ANOVA was conducted with SOA and Congruence Type as factors. Within Congruence Type levels, a repetition prime had the strongest effect on RTs. Response-congruent trials were largely facilitated compared to response-incongruent trials. Related primes facilitated responses more than unrelated primes. Thus, the main factor Congruence Type indicated a significant difference $F_{(4,28)}=37.1$, $p< .001$. The gradient of SOA shows that responses were roughly facilitated with increasing SOA. RTs changed with SOA, $F_{(6,42)}=39.8$, $p< .001$. The significant interaction [$F_{(24,168)}=4.8$, $p< .005$] explains that no differences between primes were evident at SOAs of -700 ms and -500 ms. With increasing SOA, the difference between levels of Congruence Type roughly increased.

Mean RTs of Congruence Type were illustrated in Figure 4.5 averaged over SOA. To investigate the hypotheses about effects of Semantic Relation and Response-Congruence, another ANOVA with these two factors was conducted. Semantic Relation consisted of two conditions. Trials with semantically related primes were contrasted to trials with semantically unrelated primes. Thereby, it was aggregated over responses-incongruent and response-congruent trials. Response-Congruence was calculated respectively. Congruent versus incongruent trials were aggregated over semantically related and unrelated trials.

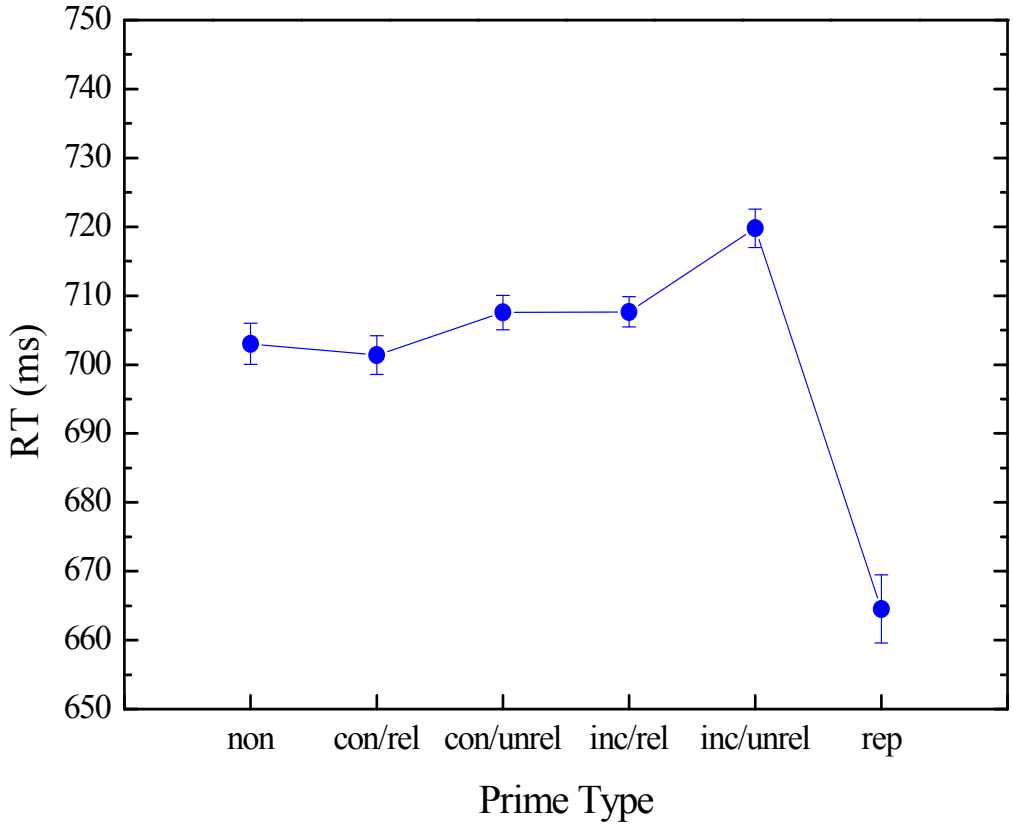


Figure 4.5: Mean RTs as a function of Congruence Type. The following Congruence Types were employed: non: no prime; con/rel: response-congruent and semantically related; con/unrel: response-congruent and semantically unrelated; inc/rel: response-incongruent and semantically related; inc/unrel: response-incongruent and semantically unrelated; rep: repetition.

Responses were facilitated by semantically related primes, as compared to unrelated primes, which was evident in a significant main effect of Semantic Relation, $F_{(1,7)}=17.0$, $p< .005$. Furthermore, Response-Congruence had a main effect, $F_{(1,7)}=10.6$, $p< .02$. This effect implied lower RTs after response-

congruent primes than after response-incongruent primes. The two factors did not interact, $F_{(1,7)}$=70.9, p= .29. Thus, Semantic Relation and Response-Congruence both affected responses independently.

4.2.3 Discussion

Effects of semantic congruence of audiovisual stimuli were once more discovered. It was shown that semantic congruence facilitated responses not just to response-congruent, but also to response-incongruent stimuli. Thus, a semantically congruent prime also had an influence when the target elicited a different response. Interestingly, effects of semantic relation were larger in response-incongruent words than in response-congruent words. One possible explanation for this result is that semantic distance may be larger for response-incongruent than for response-congruent conditions. For example, the semantically incongruent prime for the target 'horse' is 'nephew' in the response-congruent condition and 'bottle' in the response-incongruent condition. A nephew seems to be more related to a horse than a bottle (imagine a boy riding a horse in contrast to horse drinking out of a bottle). Thus, semantic distance may be a mediator for the enlarged semantic influences for response-incongruent words. Future studies may validate this by introducing semantic distance as a controlled factor.

In contrast to Experiment 9, effects were also evident at an SOA of 0 ms. Particularly, semantics had the largest influence at this SOA. The different range of SOAs might have produced this effect. The latest presentation of a prime produced the smallest semantic influences in both linguistic experiments. In Experiment 9 the lowest SOA was 0 ms, whereas in Experiment 10 it was -700 ms. Participants may calibrate their temporal concordance to the employed intervals. This hypothesis is supported by Fujisaki et al. (2004). Their participants shifted the perceived simultaneity of audiovisual stimuli to a particular time interval of an adaptation phase. This recalibration might occur in the present experiments based on the range of intervals.

The results of Experiment 9 could further be expanded because of the no-prime condition. In this condition no prime at all was presented. The advantage was that attention lay solely on the auditory stimulus, similar to the unimodal condition in the nonlinguistic experiments. Thus, no visual stimulus disturbed perception. However, SOA could therefore not be varied. No-prime condition is just comparable to the 0 ms SOA. Exploring the results at this SOA, one observes that RTs to targets without a prime lay between response-congruent and response-incongruent conditions. Response-congruence enhanced responses, whereas response-incongruence inhibited responses. This result of no primes should be tested with meaningless primes (e.g. random letters or pseudowords) in future experiments to allow comparisons at other SOAs.

Responses were furthermore affected by repetition priming. Holcomb et al. (2005) explored crossmodal linguistic repetition priming. Similar to my results, they found effects from visual primes on responses to auditory targets and vice versa. Behavioral results showed facilitated responses at all SOAs (0, 200 and 800 ms) and both directions after repeated words, compared to unrelated and pseudowords. Accordingly, ERPs revealed modulation of the N400 at most SOAs. Behavioral and ERP effects were again larger and started earlier from vision onto audition than vice versa. In contrast to the nonlinguistic experiments, presentation of the same items in both modalities was regarded as repetition instead of stimulus-congruence, because of high concordance of orthography and phonology. The large facilitation of responses after repetition primes supported this idea.

Effects of linguistic stimuli were similar to nonlinguistic stimuli. Implementing linguistic stimuli helped to support the findings that semantic congruence affects processing of audiovisual stimuli, despite the differences in stimulus properties.

5. General Discussion

The present experiments gave clear evidence of semantic influences on processing of audiovisual stimuli. Semantic congruence overcomes the difference between stimulus properties of audition and vision. Light and sound waves elicit very different sensations, but they can still be integrated to one percept. Besides spatial and temporal congruence, semantic congruence enhances integration of information from different modalities. I showed that semantic congruence affected responses despite response-congruence. In contrast to Molholm et al. (2004) and Yuval-Greenberg and Deouell (2007), semantic influences could not be explained by response-congruence. In the present non-linguistic experiments, I demonstrated that semantically congruent stimuli (e.g. sound and picture of a dog) produced faster and more accurate responses than response-congruent stimuli (e.g. horse and dog). This was confirmed especially by Experiment 1 and 2. Linguistic experiments further showed that responses to the spoken word 'brother' was facilitated by reading the word 'sister' compared to the word 'bee' (Experiment 9 and 10). Semantics also affected responses to response-incongruent stimuli, as demonstrated in Experiment 10.

Effects of semantics were found with several kinds of stimuli. Employed stimuli ranged from pictures and environmental sounds (experiments 1 through 4), over movement directions of disks and tones (experiments 5 through 8), to written and spoken words (experiments 9 and 10). Thus, results are widely generalizable. Critical points and limitations of one kind of stimuli were compensated by implementation of other stimuli. For example, pictures and sounds as well as written and spoken words were static and dynamic respectively. Some researchers might argue that these stimuli cannot elicit a coherent percept. Conversely, experiments with dynamic disks and sounds demonstrated congruence effects. On the other hand, effects of response-congruence could be ruled out in the former, but not in the latter experiments. Semantic congruence was best modifiable in words. But reading and hearing words is largely processed by same brain areas (Nakada, Fujii, Yoneoka & Kwee, 2001). The concordance of written and spoken words is based on the similarity of orthographical and phonological properties of words (Holcomb et al., 2005). Pictures and sounds instead have rather different properties. Usually we do not hear a horse, when

we see a line drawing of one. The results showed that these stimuli still can be integrated.

Stimulus-congruent and response-congruent conditions led to faster responses than unimodal stimuli. Thus, processing two stimuli was faster than processing one stimulus, although just one stimulus was needed for a correct response. This result supports that multisensory integration causes facilitation of responses. Further research is needed to prove that neural integration occurs. Miller (1982) has suggested using redundant target paradigms to show deviations from a race model to prove multisensory integration. For example, Diederich and Colonius (2004) showed that responses to trimodal and bimodal stimuli were faster than predicted by a race model. Deviations from the race model would speak against independent processing of multimodal stimuli. Such experiments require different tasks than the ones implemented here. The present study focused on semantic influences of one modality onto another. Instead a task is needed that requires processing both modalities equally. In addition, pairs of two visual stimuli, pairs of two auditory stimuli as well as pairs of one visual and one auditory stimulus have each to be presented simultaneously to allow comparisons. However, simultaneous presentation of two different sounds would result in an incomprehensible sound mix. Thus, comparisons to single unimodal stimuli were computed when possible. This comparison was regarded as sufficient to demonstrate integration, because just one modality at a time was attended to.

Responsible brain sites for the present effects would help to expand the results. PET or fMRI could explore different activations in brain areas for semantically congruent compared to semantically incongruent crossmodal stimuli. Visual semantic priming studies revealed lower activity for related compared to unrelated words in the parts of the left inferior frontal gyrus, the anterior cingulate as well as in the left superior temporal cortex (Matsumoto, Iidaka, Haneda, Okada & Sadato, 2005). Similar to results of crossmodal studies (see below), Matsumoto et al. also found a reduction of the N400 component of ERPs to related compared to unrelated words. A correlation between fMRI and ERP revealed that the source of the N400 effect is the superior temporal cortex. Further studies are needed to explore whether effects of crossmodal effects of semantic relation are located in the same regions.

Limitations of semantic influences were found by using different tasks. Semantic influences were found in a categorization task (i.e. living vs. nonliving). Tasks, which did not require processing the content, failed to reveal clear effects of semantic congruence. Such low processing tasks included detection of audiovisual stimuli (Experiment 3 and 7) and indicating which modality was presented first (Experiment 4). Semantics was completely irrelevant to fulfill these tasks. What is the reason for the absence of effects in these tasks? One possibility is that responses were overall to fast to observe different effects of the content. Responses in low level tasks were in fact faster than in the categorization tasks. For example, in Experiment 5 mean RTs lay between 450 ms and 550 ms for auditory targets and between 425 ms and 475 ms for visual targets, while RTs in Experiment 7 were only 280 ms to 290 ms. One might hypothesize that the smaller variance of mean RTs in Experiment 7 compared to Experiment 5 resulted from this floor effect. This question could be solved by implementing a study that is more difficult but does not require deeper processing. For example, stimuli, which are harder to identify, could be used.

Another possible explanation for the absence of effects of semantic relation in some of the present experiments is that in these tasks semantics was not processed. In these experiments, participants did not need to attend to the content of the stimuli to fulfill the task. Other studies have also not found effects of semantic relation. For example, Koppen et al. (2008) found no semantic influences on the Colavita effect (Colavita, 1974). Their task was to indicate presence of an auditory stimulus with one key and presence of a visual stimulus with another key. Accordingly, semantics was irrelevant and did not need to be processed. Therefore, auditory targets were detected equally often in semantically congruent trials and semantically incongruent trials. Different sizes of the Colavita effects were found for spatial and temporal congruence (Koppen & Spence, 2007). But spatial and temporal characteristics are usually processed in similar modality detection tasks (McDonald et al., 2005; Schnupp, Dawe & Pollack, 2005). Is semantic processing sufficient for effects of semantic relation? My experiments seem to agree with this point. However, effects of semantic relation have also been found with lexical decision tasks (e.g. Holcomb & Anderson, 1993). Lexical decision is widely believed to base on a purely lexical level (Grainger & Jacobs, 1996). Thus, further experiments are needed

to find out if deeper processing or semantic processing is relevant for effects of semantic relation in the present experiments. This could be achieved by exploring neuronal correlates as mentioned above.

Besides the required processing of the stimuli (low vs. deep) to fulfill the task, the level of neuronal processing is a point of interest. Literature on the early-late debate is extensive (for a review see Calvert & Thesen, 2004). Experiment 2, 9 and 10 could add results to this debate, because these experiments varied SOA. However, results were not consistent. Experiment 2 had the largest effects of semantic relation at a SOA of 200 ms for visual target and at a SOA of 350 ms for auditory targets. Simultaneity was irrelevant, because these SOAs lay outside the just noticeable difference, which resulted from Experiment 4. Strategic effects cannot solely explain the increase, because semantic influences did not increase up to the highest SOA. Strategic effects have been found to increase along with SOA (Perea & Rosa, 2002). On the other hand, participants might have had enough time to suppress the influence of the unattended stimuli at the larger SOAs. Effects of response priming have also been found to increase along with SOA (Vorberg et al., 2003), but here, category priming had no larger effects at larger SOAs (cf. Experiment 2).

Experiment 9 showed this linear increase of semantic influences along with SOA. Here, strategic effects seem to play a superior role. It seems as if the distractor needed to precede the target to receive increased effects. But Experiment 10 found the largest semantic influences at a SOA of 0 ms, which speaks against this thesis. Furthermore, strategic effects seemed not to have a large effect here, despite almost the same stimuli as in Experiment 9. To sum it up, two experiments suggest that effects occur rather late, partly because of strategic effects. Earlier processing is also supported, because semantic influences were also found at smaller SOAs.

ERP experiments could help to solve the question of processing level. Usually, effects of semantic relation are reflected by the N400 (Kutas & Federmeier, 2000). This marker is regarded to result from a rather late processing level. Would the behavioral effects in the present experiments also be found in ERPs? Molholm et al. (2004) used similar stimuli and found relatively late differences between congruent and incongruent semantics within response-

incongruent trials. In a non-hypothesized post-hoc analysis, they found a stronger negativation to incongruent stimuli in the N400. Accordingly, differences started at circa 400 ms after stimulus presentation. Yuval-Greenberg and Deouell (2007) found effects of semantic relation in gamma-band responses at a late stage. Specifically, approximately after 260 ms, responses to congruent and incongruent stimuli differed. They also searched for ERP differences, but in contrast to Molholm et al. (2004), they did not find semantic effects here. Holcomb and Anderson (1993) found a change of the N400 to linguistic semantic congruent and incongruent stimuli. Thereby, SOAs played role for visual targets. In this condition, the N400-effect was found at SOAs of 200 ms and 800 ms, but not at 0 ms. SOA had no effect on auditory targets.

Vision strongly affected audition in experiments 1, 2, 5, 6, 9, and 10. Effects from audition onto vision were overall smaller, but still evident in experiments 2, 5, and 8. Results were consistent with Yuval-Greenberg and Deouell (2007). They found stronger crossmodal semantic influences for auditory than for visual targets. What is the cause of smaller effects from audition onto vision? One possibility is that visual stimuli were easier to identify. Easier identification of visual stimuli was discovered in a preexperiment (cf. Experiment 1) and was evident by faster responses to visual targets than to auditory targets throughout all experiments. Easier identification goes along with increased reliability. More reliable information dominates multisensory perception (Welch & Warren, 1980). Furthermore, neurons integrate multisensory information in a statistical optimal fashion by considering reliability (Ernst & Banks, 2002). Thus, visual stimuli could affect auditory perception but not vice versa. This is also supported by comparing results of Experiment 1 and Experiment 2. When visual stimuli were blurred (Experiment 2), effects were slightly enlarged. Audition could influence vision, when pictures were harder to identify. However, visual stimuli were still clearly identifiable and accordingly effects were still smaller in this direction. Object identification was important for that task. Consequently, vision should dominate over audition in such a task. Another effect of better identification of visual stimuli is that responses to visual stimuli were simply too fast to find differences. A floor effect resulted. Furthermore, usually

humans identify objects because of their visual impression instead of the auditory properties. A more difficult task could help to solve this hypothesis.
Experiment 8 specifically focused on effects from audition onto vision and found explicit evidence for semantic influences. The path of the disks was ambiguous in one condition. Strong effects of the auditory stimuli were found. However, when the path was defined by different colored disks, the sounds also affected responses to the visual stimuli. Does this imply that reliability is not important? Not really, because the auditory stimulus probably primed the response to the motion path. The sound was presented at the coincidence. At this time, the path was not clearly identifiable. On the other hand, no effect of sounds on errors was found and error rates were small. Thus, the visual stimuli were sufficiently reliable for responding correctly.

A remaining open question is how semantic congruence emerges in humans. A preexperiment tried to discover effects of learned congruence between sounds with different pitch and perceived path of motion (based on Experiment 8). After approximately 500 learning trials, the learned congruence had no effects. Furthermore, blockwise analyses of the current experiments revealed no enlarged effects after multiple repetitions of the stimulus combinations. Thus, it is believed that the semantics of the employed stimuli was naturally learned during childhood. One example of crossmodal effects in infants used the ambiguous stream-bounce display as visual stimuli (Scheier, Lewkowicz & Shimojo, 2003). Responses of 6- and 8-moths old infants were different for presenting a sound at the coincidence compared to an unsynchronized sound. 4-months old infants did not show this difference. This effect is not caused by semantics but rather by attention and knowledge of objects. Crossmodal effects of semantic relation were explored in infants by Friedrich and Friederici (2004). They presented pictures and congruent or incongruent spoken words to 19-months-old children while measuring ERPs. The typical N400 modulation was found. Thus, the ability of integrating semantic information is present already at young age.
A further question is what if one sense is absent? Blind and deaf persons cannot profit from audiovisual semantic congruence. Are they instead using intramodal semantic congruence? Röder, Demuth, Streb and Rösler (2003) ad-

dressed this question in an auditory priming study. They found no difference between congenially blind participants and sighted controls in the size of the semantic priming effect. They found generally faster responses to all targets (words and pseudowords) in blind compared to sighted participants. Thus, faster language processing was evident in blinds, but enhancement of semantic congruence was not more than in sighted people.

What affects semantic influences on multisensory integration? To sum it up, the present experiments demonstrated that the task has a strong influence on whether effects of semantic relation are present or not. Processing of the content was important for effects of semantics. Temporal aspects differed for the present experiments. Overall, stronger semantic influences were found when the distractor preceded the target. Thus, semantic influences were found at a rather late processing stage. Perceived simultaneity of the multisensory events was not essential. Increased effects of semantic relation were found in classification tasks. Effects from vision onto audition dominate.

Crossmodal semantic influences are incessantly used in our daily lives. Returning to the introducing example of the zoo visit, how does our brain process seeing an elephant and hearing a lion? Based on the current experiments, we know that the brain takes longer to process incongruent than congruent crossmodal information (i.e. a roaring elephant vs. a trumpeting elephant). It seems that our neurons are as confused as we are when perceiving a roaring elephant.

6. References

Alais, D. & Burr, D. (2004). No direction-specific bimodal facilitation for audiovisual motion detection. *Cognitive Brain Research, 19*, 2, 185-194.

Alvarado, J.C., Vaughan, J.W., Stanford, T.R. & Stein, B.E. (2007). Multisensory versus unisensory integration: contrasting modes in the superior colliculus. *Journal of Neurophysiology, 97,* 3193-3205.

Arnold, D. H., Johnston, A. & Nishida, S. (2005). Timing sight and sound. *Vision Research, 45*, 1275-1284.

Arrighi, R., Alais, D., Burr, D. (2004). Perceiving timing of first- and second-order changes in vision and hearing. *Experimental Brain Research, 166*, 445-454.

Arrighi, R., Alais, D., Burr, D. (2006). Perceptual synchrony of audiovisual streams for natural and artificial motion sequences, *Journal of Vision, 6,* 260-268.

Banks, M.S. (2004). What you see and hear is what you get. *Current Biology, 14*, 1-3.

Barbarotto, R., Laiacona, M., Macchi, V. & Capitani, E. (2002). Picture reality decision, semantic categories and gender. A new set of pictures, with norms and an experimental study. *Neuropsychologia, 40*, 1637-1653.

Barraclough, N.E., Xiao, D., Baker, C.I., Oram, M.W. & Perrett, D.I. (2005). Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *Journal of Cognitive Neuroscience, 17*, 3, 377-391.

Bologini, N., Rasi, F., Làdavas, E. (2005). Visual localization of sounds. *Neuropsychologia, 43*, 1655-1661.

Brainard, D.H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433-436.

Bremmer, F., Schlack, A., Shah, N.J., Zafiris, O., Kubischik, M., Hoffmann, K., Zilles, K. & Fink, G.R. (2001). Polymodal motion processing in posterior parietal and premotor cortex: a human fMRI study strongly implies equivalencies between humans and monkeys. *Neuron, 29*, 1, 287-296.

Bushara, K., Hanakawa, T., Immisch, I., Toma, K., Kansaku, K. & Hallett, M. (2002). Neural correlates of cross-modal binding. N*ature Neuroscience 6,* 2, 190-195.

Calvert, G.A. (2001). Crossmodal Processing in the Human Brain: Insights from functional neuroimaging studies. *Cerebral Cortex, 11*, 1110-1123.

Calvert, G.A., Brammer, M.J., Bullmore, E.T., Campbell, R., Iversen, S.D. & David, A.S. (1999). Response amplification in sensory-specific cortices during crossmodal binding. *NeuroReport, 10*, 2619-2623.

Calvert, G.A., Brammer, M.J. & Iversen, S.D. (1998). Crossmodal identification. *Trends in Cognitive Sciences, 2*, 7, 247-253.

Calvert, G.A., Bullmore, E.T., Brammer, M.J., Campbell, R., William, S.C., Mcguire, P.K., Woodruff, P.W., Iversen, S.D., David, A.S. (1997). Activation of auditory cortex during silent lipreading. *Science, 276*, 593-596.

Calvert, G.A., Campbell, R. & Brammer, M.J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology, 10*, 649-657.

Calvert, G.A. & Lewis, J.W. (2004). Hemodynamic studies of audio-visual interactions. Chapter 30, 483-502. In: Calvert, G.A., Spence, C. & Stein, B.E. (Eds.) (2004). *The handbook of multisensory processing*. Cambridge, USA: MIT Press.

Calvert, G.A. & Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. *Journal of Physiology, 98*, 191-205.

Campbell, R. (1998). How brains see speech: the cortical localization of speechreading in hearing people. Chapter 9, 177-193. In: Campbell, B., Dodd, B. & Burnham, D. (Eds.). *Hearing by eye II: Advances in the psychology of speechreading and auditory-visual speech.* Hove, UK: Taylor & Francis.

Colavita, F.B. (1974). Human sensory dominance. *Perception & Psychophysics, 16*, 2, 409-412.

Cytowic, R.E. (2002). *Synesthesia. A union of the senses. 2nd Ed*. Cambridge: MIT Press.

Desimone, R. & Gross, C.G. (1979). Visual areas in the temporal cortex of the macaque. *Brain Research, 178*, 363-380.

Desimone, R. & Ungerleider, L.G. (1986). Multiple visual areas in the caudal superior temporal sulcus of the macaque. *Journal of Comparative Neurology, 248*, 2, 164-189.

Diderich & Colonius (2004) Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time. *Perception & Psychophysics, 66*, 8, 1388-1404.

Driver, J. & Spence, C. (2000). Multisensory perception: beyond modularity and convergence. *Current Biology, 10*, 731-735.

Encyclopedia Britannica (2002). The new Enyclopedia Britannica. Chicago: Encyclopedia Britannica.

Ernst, M.O. & Banks, M.S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*, 429-433.

Ettlinger, G. & Wilson, W.A. (1990). Coss-modal performance. behavioural processes, phylogenetic considerations and neural mechanisms. *Behavioural Brain Research, 40*, 3, 169-192.

Fink, M., Ulbrich, P., Churan, J. & Wittmann, M. (2005). Stimulus-dependent processing of temporal order. *Behavioural Processes, 71*, 344-352.

Foxe, J.J. & Schroeder, C.E. (2005). The case for feedforward multisensory convergence during early cortical processing. *NeuroReport, 16*, 5, 419-423.

Friedrich, M. & Friederici, A.D. (2004). N400-like semantic incongruity effect in 19-month-olds: processing known words in picture contexts. *Journal of Cognitive Neuroscience, 16*, 1465-1477.

Fujisaki, W., Shimojo, S., Kashino, M. & Nishida, S. (2004). Recalibration of audiovisual simultaneity. N*ature Neuroscience, 7*, 7, 773-778.

Fuster, J.M., Bodner, M. & Kroger, J.K. (2000). Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature, 405*, 347-351.

Giard, M.H. & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *Journal of Cognitive Neuroscience, 11*, 5, 473-490.

Gondan, M. (2005). Multisensorische Integration von redundanten Reizen. *Dissertation. Philipps-Universität Marburg*.

Gottfried, J.A. & Dolan, R.J. (2003). The nose smells what the eye sees: crossmodal visual facilitation of human olfactory perception. *Neuron, 39*, 375-386.

Gray, R., Tan, H.Z. & Young, J.J. (2002). Do multimodal signals need to come from the same place? Crossmodal attentional links between proximal and

distal surfaces. *Fourth IEEE International Conference on Multimodal Interfaces*, 437.

Hauser, M.D., Chomsky, N. & Fitch, W.T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science, 298*, 1569-1579.

Helbig, H.B. & Ernst, M.O. (2007). Optimal integration of shape information from vision and touch. *Experimental Brain Research, 179*, 595-606.

Heron, J., Whitaker, D., McGraw, P.V. (2004). Sensory uncertainty governs the extent of audio-visual interaction. *Vision Research, 44*, 2875-2884.

Holcomb, P.J. & Anderson, J.E. (1993). Cross-modal semantic priming: a time-course analysis using event-related brain potentials. *Language and Cognitive Processes, 8*, 4, 379 – 411.

Holcomb, P.J., Anderson, J.E. & Grainger, J. (2005). An electrophysiological study of cross-modal repetition priming. *Psychopysiology, 42*, 493-507.

Holcomb, P.J. & Neville, H.J. (1990). Auditory and visual semantic priming in lexical decision: a comparison using event-related brain potentials. *Language and Cognitive Processes, 5*, 4, 281-312.

Holmes, N.P. & Spence, C. (2005). Multisensory integration: space, time and superadditivity. *Current Biology*, *15*, 18, 762-764.

Howard, & Tempelton, (1966). *Human spatial orientation*. London: Wiley.

Jáskowski, P., Jaroszyk, F. & Hojan-Jezierska, D. (1990). Temporal-order judgments and reaction time for stimuli of different modalities. *Psychological Research, 52*, 35-38.

Johnson, J.A. & Zatorre, R.J. (2005). Attention to simultaneous unrelated auditory and visual events: behavioral and neural correlates. *Cerebral Cortex, 15,* 1609-1620.

Kawachi, Y. & Gyoba, J. (2006). Presentation of a visual nearby moving object alters stream/ bounce event perception. *Perception, 35*, 1289-1294.

Keetels, M., & Vroomen, J. (2005).The role of spatial disparity and hemifields in audio-visual temporal order judgments. *Experimental Brain Research, 167,* 635-640.

Kirchner & Colonius (2005). Cognitive control can modulate intersensory facilitation: speeding up visual antisaccades with an auditory distractor. *Experimental Brain Research, 166*, 440-444.

Kitagawa, N. & Ichihara, S. (2002). Hearing visual motion in depth. *Nature, 416,* 172-174.

Koppen, C., Alsius, A. & Spence, C. (2008). Semantic congruency and the colavita visual dominance effect. *Experimental Brain Research, 184*,533-546.

Koppen, C. & Spence, C. (2007). Seeing the light: exploring the Colavita visual dominance effect. *Experimental Brain Research, 180*, 737-754.

Kutas, M. & Federmeier, K.D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Science, 4*, 463-470.

Laurienti, P.J., Kraft, R.A., Maldjian, J.A., Burdette, J.H. & Wallace, M.T. (2004). Semantic congruence is a critical factor in multisensory behavioural performance. *Experimental Brain Research, 158*, 405-414.

Lehmann, S. & Murray, M.M. (2005). The role of multisensory memories in unisensory object discrimination. *Cognitive Brain Research, 24*, 326-334.

Loftus, G.R. & Masson, M.E.J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1,* 476-490.

Macaluso, E., Frith, C.D. & Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science, 289*, 1206-1208.

Maeda, F. Kanai, R. & Shimojo, S. (2004). Changing pitch induced visual motion illusion. *Current Biology, 14,* 23, 990-991.

McDonald, J.J., Teder- Sälejärvi, W.A. & Hillyard, S.A. (2000). Involuntary orienting to sound improves visual perception. *Nature, 407*, 906-908.

McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746-748.

Meredith, M.A., (2002). On the neuronal basis for multisensory convergence: a brief overview. *Cognitive Brain research, 14*, 31-40.

Meredith, M.A. & Stein, B.E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science, 221*, 389-391.

Meredith, M.A. & Stein, B.E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology, 56*, 3, 640-662.

Meredith, M.A. & Stein, B.E. (1996). Spatial determinants of multisensory integration in cat superior colliculus neurons. *Journal of Neurophysiology, 75*, 5, 1843-1857.

Metzger, W. (1934). Beobachtungen über phänomenale Identität. *Psychologische Forschung*, 19, 1-60.

Miller, J. (1982). Divided attention: evidence for coacivation with redundant signals. *Cognitive Psychology, 14*, 247-279.

Molholm, S., Ritter, W., Javitt, D.C. & Foxe, J.J. (2004). Multisensory visual-auditory object recognition in humans: a high-density electrical mapping study. *Cerebral Cortex, 14,* 452-465.

Morton, J., Marcus, S.M. & Frankish, C. (1976). Perceptual centers (P-centers). *Psychological Review, 83*, 405-408.

Munhall, K.G. & Vatikiotis-Bateson, E. (1998). The moving face during speech communication. Chapter 6, 123-139. In: Campbell, B., Dodd, B. & Burnham, D. (Eds.). *Hearing by eye II: Advances in the psychology of speechreading and auditory-visual speech.* Hove, UK: Taylor & Francis.

Murray, M.M., Foxe, J.J. & Wylie, G.R. (2005). The brain uses single-trial multisensory memories to discriminate without awareness. *NeuroImage, 27*, 473-478.

Murray, M.M., Molholm, S., Michel, C.M., Heslenfeld, D.J., Ritter, W., Javitt, D.C., Schroeder, C.E. & Foxe, J.J. (2005). Grabbing your ear: rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. *Cerebral Cortex, 15*, 963-974.

Musacchia, G., Sams, M., Nicol, T. & Kraus, N. (2006). Seeing speech affects acoustic information processing in the human brainstem. *Experimental Brain Research, 168*, 1-10.

Nagy, A., Eördegh, G., Paróczy, Z., Márkus, Z. & Benedek, G. (2006). Multisensory integration in the basal ganglia. *European Journal of Neuroscience, 24*, 917-924.

Nakada, T., Fujii, Y., Yoneoka, Y. & Kwee, I.L. (2001). Planum temporale: where spoken and written language meet. *European Neurology, 48*, 121-125.

Navarra, J., Vatakis, A., Zampini, M., Soto-Faraco, S., Humphreys, W. & Spence, C. (2005). Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration. *Cognitive Brain Research, 25*, 499-507.

Neuhoff, J.G. (1998). Perceptual bias for rising tones. *Nature, 395*, 123.

Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437-442.

Perea, M. & Rosa, E. (2002). Does the proportion of associatively related pairs modulate the associative priming effect at very brief stimulus-onset asynchronies? *Acta Psychologica, 110*, 103-124.

Perea, M., Rosa, E. & Gómez, C. (2002). Is the go/no-go lexical decision task an alternative to the yes/no lexical decision task? *Memory & Cognition, 30*, 1, 34-45.

Ptito, M., Moesgaard, S.M., Gjedde, A. & Kupers, R. (2005). Cross-modal plasticity revealed by electrotactile stimulation of the tongue in the congenitally blind. *Brain, 128*, 606-614.

Purves, D., Augustine, G.J., Fitzpatrick, D., Hall, W.C., Lamantia, A.-S., McNamara, J.O. & Williams, S.M. (2004). *Neuroscience, 3rd Ed*. Sunderland, MA: Sinauer.

Quillian, M.R. (1962). A revised design for an understanding machine. *Mechanical Translation, 7*, 17-29.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*, 3, 510-532.

Remjin, G.B., Ito, H. & Nakajima, Y. (2004). Audiovisual integration: an investigation of the 'streaming-bouncing' phenomenon. *Journal of Physiological Anthropology and Applied Human Science, 23,* 243-247.

Reynvoet, B., Brysbaert, M. & Fias, W. (2002). Semantic priming in number naming. *The Quarterly Journal of Experimental Psychology, 55A*, 4, 1127-1139.

Röder, B., Demuth, L., Streb, J. & Rösler, F. (2003). Semantic and morpho-syntactic priming in auditory word recognition in congenitally blind adults. *Language and Cognitive Processes, 18*, 1, 1-20.

Rosenblum, L.D. & Saldaña, H.M. (1996). An audiovisual test of kinematic primitives for visual speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 2, 318-331.

Rossion, B. & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception, 33*, 217-236.

Roufs, J.A.J. (1974). Dynamic properties of vision, V: Perception lag and reaction time in relation to flicker and flash threshold. *Vision Research, 14*, 853-869.

Sadato, N., Pascual-Leone, A., Grafman, J., Ibanez, V., Deiber, M.P., Dold, G. & Hallett, M. (1996). Activation of the primary visual cortex by Braille reading in blind subjects. *Nature, 380*, 526-528.

Sanabria, D., Correa, Á, Lupiáñez. & Spence, C. (2004). Bouncing or streaming? Exploring the influence of auditory cues on the interpretation of ambiguous visual motion. *Experimental Brain Research, 157*, 537-541.

Scheier, C., Lewkowicz, D.J. & Shimojo, S. (2003). Sound induced perceptual reorganization of an ambiguous motion display in human infants. *Developmental Science, 6*, 3, 233-244.

Schnupp, J.W.H., Dawe, K.L., Pollack, G.L. (2005). The detection of multisensory stimuli in an orthogonal sensory space. *Experimental Brain Research, 162*, 2, 181-190 .

Scott, S.K. (1998). The point of P-centres. *Psychological Research, 61*, 4-11.

Sekuler, A.B. & Sekuler, R. (1999). Collisions between moving visual targets: what controls alternative ways of seeing an ambiguous display? *Perception, 28*, 415-432.

Sekuler, R., Sekuler, A.B. & Lau, R. (1997). Sound alters visual motion perception. *Nature*, *385*, 308.

Shams, L., Kamitani, Y. & Shimojo, S. (2004). Modulations of visual perception by sound. Chapter 2, 27-33. In: Calvert, G.A., Spence, C. & Stein, B.E. (Eds.) (2004). *The handbook of multisensory processing*. Cambridge, USA: MIT Press.

Shams, L., Kamitani, Y., Thompson, S. & Shimojo, S. (2001). Sound alters visual evoked potentials in humans. *Cognitive Neuroscience and Neuropsychology, 12*, 174, 3849-3852.

Shimojo, S. & Shams, L. (2001). Sensory modalities are not separate modalities: plasticity and interactions. *Current Opinion in Neurobiology, 11,* 505-509.

Sinnett, S., Juncadella, M., Rafal, R., Azañón, E. & Soto-Faraco, S. (2007). A dissociation between visual and auditory hemi-inattention: Evidence from temporal order judgements. *Neuropsychologia, 45,* 552-560.

Snodgrass, J.G. & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology, 6*, 2, 174-215.

Spence, C. & Driver, J. (2004). *Crossmodal space and crossmodal attention*. Oxford: Oxford University Press.

Stanford, T.R., Quessy, S. & Stein, B.E. (2005). Evaluating the operations underlying multisensory integration in the cat superior colliculus. *The Journal of Neuroscience, 25*, 28, 6499-6508.

Stein, B.E, Jiang, W. & Stanford, T.R. (2004). Multisensory integration in single neurons of the midbrain. Chapter 15, 243-264. In: Calvert, G.A., Spence, C. & Stein, B.E. (Eds.) (2004). *The handbook of multisensory processing*. Cambridge, USA: MIT Press.

Stein, B.E. & Meredith, M.A. (1993). *The merging of the senses*. Cambridge, USA: MIT Press.

Stein, B.E., Meredith, M.A., Huneycutt, W.S. & McDade, L. (1989). Behavioral indices of multisensory integration: Orientation to visual cues is affected by auditory stimuli. *Journal of Cognitive Neuroscience, 1*, 12-24.

Sugita, Y. & Suzuki, Y. (2003). Implicit estimation of sound-arrival time. N*ature, 421*, 911.

Sumby, W.H. & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America, 26*, 2, 212-215.

Sur, M. (2004). Rewiring cortex: Cross-modal plasticity and its implications for cortical development and function. Chapter 42, 681-694. In: Calvert, G.A., Spence, C. & Stein, B.E. (Eds.) (2004). *The handbook of multisensory processing*. Cambridge, USA: MIT Press.

Tervaniemi, M. & Hugdahl, K. (2003). Lateralization of auditory-cortex functions. *Brain Research Reviews, 43*, 231-246.

Todd, J.W. (1912). *Reaction to multiple stimuli*. New York: Science Press.

Tuomainen, J., Andersen, T.S., Tiippana, K. & Sams, M. (2005). Audio-visual speech perception is special. *Cognition, 96*, B13-22.

Ungerleider, L.G. & Mishkin, M. (1982). Two cortical visual systems. In: D.J. Ingle, D.J., Goodale, M.A. & Mansfield, R.J.W. (Eds). *Analysis of Visual Behavior.* Cambridge, MA: MIT, 549-586.

Vorberg, D., Mattler, U., Heinecke, A., Schmidt, T. & Schwarzbach, J. (2003). Different time courses for visual perception and action priming. *Proceedings of the National Academy of Sciences, USA, 100*, 6275-6280.

Wallace, M.T. (2004). The development of multisensory processes. *Cognitive Processing, 5*, 2, 69-83.

Wallace, M.T., Wilkinson, L.K. & Stein, B.E. (1996). Representation and integration of multiple sensory inputs in primate superior colliculus. *Journal of Neurophysiology, 76*, 2, 1246-1266.

Watanabe, M. (1992). Frontal units of the monkey coding the associative significance of visual and auditory stimuli. *Experimental Brain Research, 89*, 2, 233-247.

Watanabe, K. & Iwai, E. (1991). Neuronal activity in visual, auditory and polysensory areas in the monkey temporal cortex during visual fixation task. *Brain Research Bulletin, 26*, 4, 583-592.

Watanabe, K. & Shimojo, S. (1998). Attentional modulation of visual motion events. *Perception, 27,* 9, 1041-1054.

Watanabe, K. & Shimojo, S. (2001a). When sound affects vision: Effects of auditory grouping on visual motion perception. *Psychological Science, 12*, 2, 109-116.

Watanabe, K. & Shimojo, S. (2001b). Postcoincidence trajectory duration affects motion event perception. *Perception & Psychophysics, 63,* 1, 16-28.

Welch, R.B. & Warren, D.H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88*, 3, 638-667.

Wilcox, R.R. (1993). Analysing repeated measures or randomized block design using trimmed means. *British Journal of Mathematical and Statistical Psychology, 46*, 63-76.

Woodworth, R.S. & Schlosberg, H. (1954). *Experimental Psychology*. Revised Edition. New York: Holt, Rinehart & Winston.

Yuval-Greenberg, S. & Deouell, L.Y. (2007). What you see is not (always) what you hear: induced gamma band responses reflect cross-modal interactions in familiar object recognition. *The Journal of Neuroscience, 27,* 5, 1090-1096.

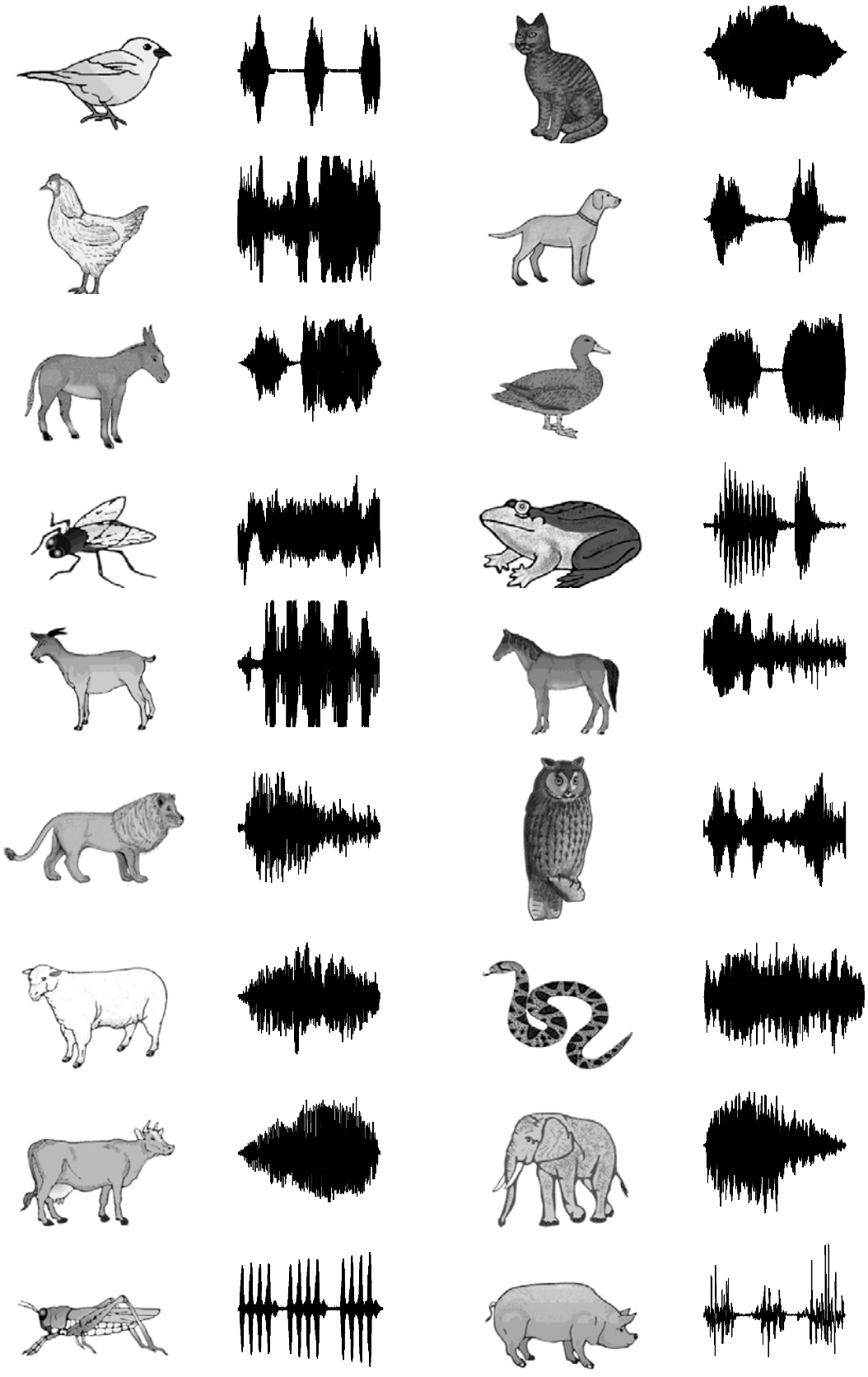
Zampini, M., Brown, T., Shore, D.I., Maravita, A., Röder, B. & Spence, C. (2005). Audiotactile temporal order judgments. *Acta Psychologica, 118*, 277-291.

Zampini, M., Guest, S., Shore, D.I. & Spence, C. (2005). Audio-visual simultaneity judgments. *Perception & Psychophysics*, *67,* 3, 531-544.

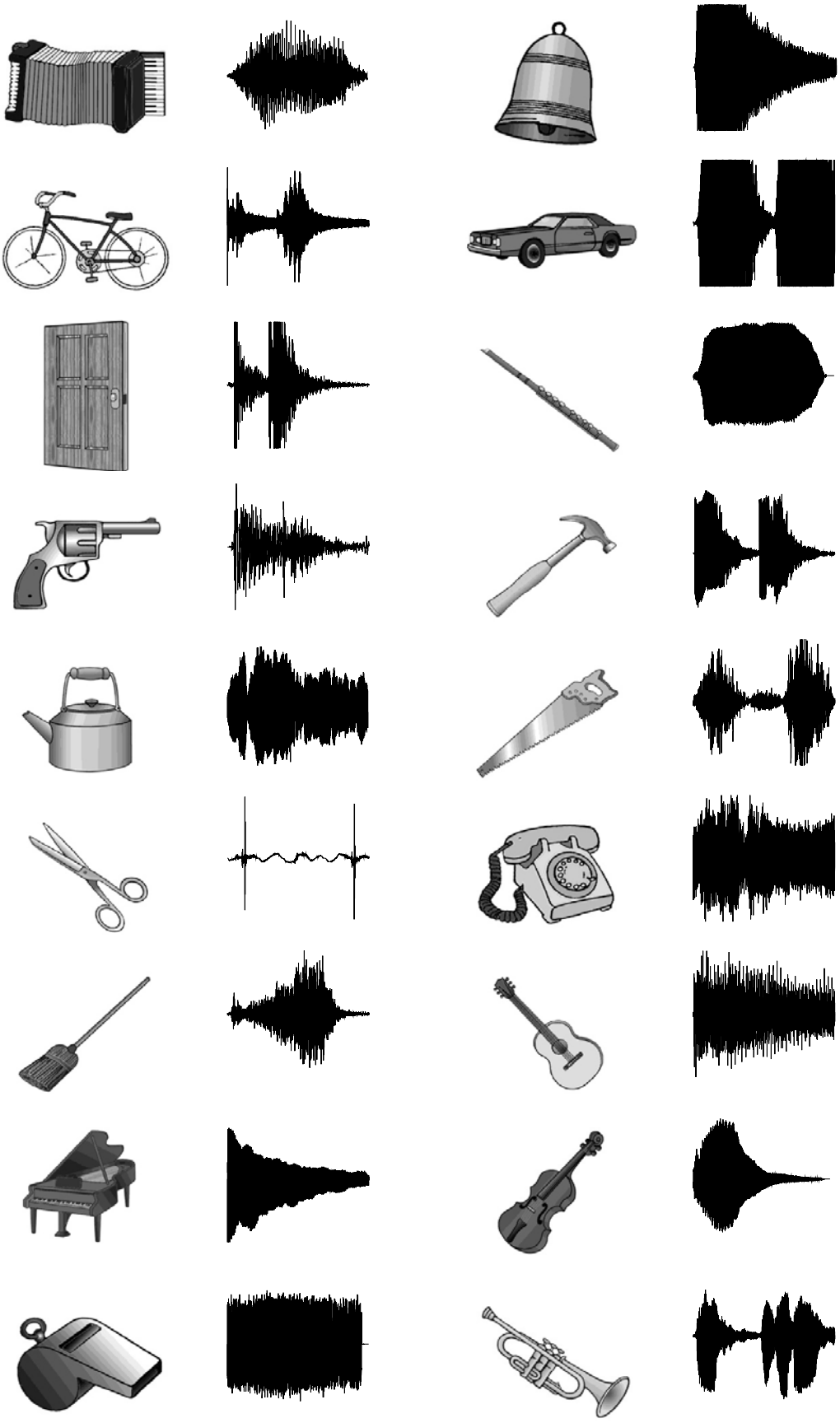
Zampini, M., Shore, D.I. & Spence, C. (2003). Audiovisual temporal order judgments. *Experimental Brain Research, 152*, 198-210.

7. Appendix

Visual and waveform of auditory stimuli categorized as living of Experiment 1



Visual and waveform of auditory stimuli categorized as nonliving of Experiment 1



Prime- and target-words of Experiment 9

Target (auditory)	Primes (visual)		
	response-congruent / related	response-congruent / unrelated	response-incongruent / unrelated
Bruder	Schwester	Hummel	Fernseher
Arzt	Doktor	Affe	Ruder
Onkel	Tante	Floh	Auto
Räuber	Dieb	Lurch	Motor
Vater	Mutter	Kalb	Säge
Richter	Anwalt	Schwein	Ampel
Jäger	Förster	Rind	Ball
Schlange	Natter	Freund	Straße
Pferd	Pony	Neffe	Hof
Gans	Ente	Kind	Treppe
Biene	Wespe	Bauer	Kabel
Frosch	Kröte	König	Dose
Tiger	Löwe	Zwerg	Rohr
Kamel	Lama	Frau	Geige
Herz	Lunge	Schaf	Stecker
Nase	Mund	Held	Flöte
Ohr	Auge	Bambi	Licht
Darm	Magen	Mann	Antenne
Vene	Arterie	Käfer	Sieb
Finger	Hand	Mädchen	Sessel
Leber	Milz	Pfau	Seil
Hammer	Nagel	Soße	Lehrer
Messer	Gabel	Grütze	Troll
Teller	Tasse	Sekt	Mörder
Tisch	Stuhl	Toast	Fliege
Papier	Blatt	Wasser	Hund
Foto	Bild	Honig	Maus
Tür	Fenster	Nudel	Huhn
Bier	Wein	Knopf	Katze
Pfeffer	Salz	Stiefel	Ratte
Saft	Nektar	Rock	Tänzer
Brot	Butter	Ring	Schüler
Bonbon	Zucker	Robe	Vogel
Torte	Kuchen	Ärmel	Hengst
Käse	Wurst	Kleid	Sieger
Hut	Kappe	Bohrer	Rabe
Schuh	Socken	Säge	Solist
Hemd	Hose	Kanne	Raupe
Jacke	Mantel	Lampe	Stute
Schal	Tuch	Krug	Wurm
Kette	Brosche	Topf	Turner
Riemen	Gürtel	Beil	Elster

Prime- and target-words of Experiment 10

Target (auditory)	Primes (visual) response-congruent		response-incongruent	
	related	**unrelated**	**related**	**unrelated**
Bruder	Schwester	Hummel	Kloster	Fernseher
Arzt	Doktor	Affe	Praxis	Ruder
Räuber	Dieb	Lurch	Bank	Motor
Vater	Mutter	Kalb	Gebet	Pfeil
Richter	Anwalt	Schwein	Robe	Ampel
Jäger	Förster	Qualle	Flinte	Ball
Schlange	Natter	Freund	Gift	Straße
Pferd	Pony	Neffe	Sattel	Flasche
Gans	Ente	Opa	Feder	Treppe
Biene	Wespe	Bauer	Honig	Kabel
Tiger	Löwe	Zwerg	Käfig	Rohr
Kamel	Lama	Greis	Wüste	Geige
Nase	Mund	Held	Duft	Flöte
Ohr	Auge	Reh	Musik	Wolke
Darm	Magen	Mann	Klo	Antenne
Vene	Arterie	Käfer	Spritze	Sieb
Finger	Hand	Mädchen	Ring	Sessel
Leber	Milz	Pfau	Alkohol	Seil
Hammer	Nagel	Soße	Schmied	Lehrer
Messer	Gabel	Grütze	Mörder	Troll
Tisch	Stuhl	Milch	Kellner	Fliege
Papier	Blatt	Wasser	Dichter	Hund
Foto	Bild	Senf	Modell	Bär
Tür	Fenster	Nudel	Tischler	Huhn
Bier	Wein	Knopf	Säufer	Katze
Pfeffer	Salz	Stiefel	Koch	Ratte
Brot	Butter	Ring	Bäcker	Tänzer
Bonbon	Zucker	Anzug	Kind	Vogel
Torte	Kuchen	Ärmel	Konditor	Hengst
Käse	Wurst	Kleid	Maus	Sieger
Hut	Kappe	Bohrer	Kopf	Rabe
Schuh	Socke	Säge	Fuß	Solist
Hemd	Hose	Kanne	Brust	Raupe
Jacke	Mantel	Lampe	Schneider	Stute
Schal	Tuch	Krug	Hals	Wurm
Kette	Brosche	Topf	Hund	Turner